

10/10/2003

GenCore version 5.1.6  
 Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 8, 2003, 16:19:57 ; Search time 46.2838 Seconds

(without alignments)  
 4811.606 Million cell updates/secTitle: US-10-002-309B-2\_COPY\_24\_886  
 Perfect score: 4597

Sequence: 1 ADNNSAIYFTNTSQPINDLQG.....AWNGOYLDLFSKPRSMRVTYK 863

Scoring table: BLASTM62

Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

SPTREMBL\_23.1:

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_Plant:\*
- 11: sp\_rabbit:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_archeap:\*
- 17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	4597	100.0	898	2	082882
2	1436	31.0	2	Q9ZAL1	082882 escherichia
3	686.5	14.9	1002	2	Q9AGX6
4	671	14.6	1335	16	Q9KN18
5	571	12.4	875	16	Q8D5P4
6	290	6.3	957	16	Q9KPH2
7	166.5	3.6	1217	16	Q9TGH3
8	161	3.5	1038	2	Q45780
9	159.5	3.5	2367	2	Q46034
10	158.5	3.4	204	5	Q8IL70
11	156.5	3.4	1157	16	Q9TF62
12	154.5	3.4	2367	2	Q9F931
13	149	3.2	1284	2	Q8RQ09
14	148.5	3.2	2468	16	Q91243
15	148.5	3.2	3692	2	Q8KRR3
16	148	3.2	1021	2	Q46085

ALIGNMENTS			
RESULT 1			
082882	PRELIMINARY;	PRT;	898 AA.
ID 082882; Q9ZG01;			
AC 082882;			
DT 01-NOV-1998 (TREMBLrel. 08, Created)			
DR 01-NOV-1999 (TREMBLrel. 12, Last sequence update)			
DR 01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE Plasmid PO157 DNA, complete sequence.			
GN L7031.			
OS Escherichia coli O157:H7.			
Plasmid PO157.			
OG CC			
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OX NCBI_TAXID=803334;			
RN [1]			
SEQUENCE FROM N.A.			
STRAIN=O157:H7;			
RC MEDLINE=98280540; PubMed=9628576;			
RA Makino K., Ishii K., Yokoyama K., Hattori M., Yamasaki T., Iida T., Yamamoto K., Honda T., Yutsudo H.C., Kubota Y., Yamauchi Y., Han C.G., Ohnishi E., Kasanatsu M., Hayashi T., Kuhara S., Shinagawa H.;			
RA [2]			
RT "Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic Escherichia coli O157:H7 derived from Sakai outbreak.";			
RC STRAIN=EDL933;			
RA Burland V., Shao Y., Perna N.T., Plunkett G., Sofia H.J., Blattner F.R.;			
RA [3]			
RT "The complete DNA sequence and analysis of the large virulence plasmid of Escherichia coli O157:H7."			
RC Nucleic Acids Res. 26:4196-4204(1998).			
RA [4]			
RP SEQUENCE FROM N.A.			
RC STRAIN=EDL933;			
RA Brundin W.;			

BL	Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
DR	EMBL; AB011549; BAA31757.3; -;
DR	EMBL; AF074613; AAC7099.1; -;
DR	EMBL; Y11831; CAA7251.1; -;
KN	Hypothetical protein; Plasmid.
SEQUENCE	898 AA; <a href="#">9048</a> ; 3C1AE23E3EA1FAB CRC64;
Query Match	100.0%; Score: 4597; DB: 2; Length: 898;
Best Local Similarity	100.0%; Pred. No. 3e-294; Mismatches: 0;保守性: 0;
Matches	863; Conservative: 0; Mismatches: 0; Indels: 0; Gaps: 0;
Qy	1 ADNNSAIYFTNTSQINDQSLAAEVKAQSIQLPAHPKGDSOPHTLSRKSLLLVRPV 60
Db	36 ADNNSAIYFTNTSQINDQSLAAEVKAQSIQLPAHPKGDSOPHTLSRKSLLYVRPV 95
Qy	61 KADDKTPVQEYERDDNNKILGTLTYPSSLPDITYHLDGVPEEGIDTPHNNTKKINT 120
Db	96 KADDKTPVQEYERDDNNKILGTLTYPSSLPDITYHLDGVPEEGIDTPHNNTKKINT 155
Qy	121 VAEVNLTSDASCSSTISHLNNALVEHTANGRWRDYLPLQGPDLLEGKMRVFRSSAGYS 180
Db	156 VAEVNKLSDASCSSTISHLNNALVEHTANGRWRDYLPLQGPDLLEGKMRVFRSSAGYS 215
Qy	181 STVFYGDRKVTLISVGNTLFLPKVNGWFRSSELLENRRNITYAQH1WSAELPAHNTVPGNL 240
Db	216 STVFYGDRKVTLISVGNTLFLPKVNGWFRSSELLENRRNITYAQH1WSAELPAHNTVPGNL 275
Qy	241 VIKQGNLSGRNDIKIGAPGELLATIDGMITPDRDFAKKEAHKEYFTOTIPVSRM 300
Db	276 VIKQGNLSGRNDIKIGAPGELLATIDGMITPDRDFAKKEAHKEYFTOTIPVSRM 335
Qy	301 IYNNAAPHLKEVMPPTGELLTMDPGNGKHSQPMQRICKELVSHGLDNANGLNSTA 360
Db	336 IYNNAAPHLKEVMPPTGELLTMDPGNGKHSQPMQRICKELVSHGLDNANGLNSTA 395
Qy	361 GLGENSHPPVVAQLAHHNSRGNYANGIQVHRGSGGGGIVTLDSTLGNEFSHEVCHNYGLG 420
Db	396 GLGENSHPPVVAQLAHHNSRGNYANGIQVHRGSGGGGIVTLDSTLGNEFSHEVCHNYGLG 455
Qy	421 HYDGFIFGSVTRSAEENNNTGWDGDKKRFTPNFTPSQTTEKSCINNQOCEPFQHGKRF 480
Db	456 HYDGFIFGSVTRSAEENNNTGWDGDKKRFTPNFTPSQTNEKSCINNQOCEPFQHGKRF 515
Qy	481 DAMAGGSPEPSAANRFMTPNSAITORFFENKAYEDRSRSTGSKWNADEQMEPYEHT 540
Db	516 DAMAGGSPEPSAANRFMTPNSAITORFFENKAYEDRSRSTGSKWNADEQMEPYEHT 575
Qy	541 IDRAEQITASVNLSESKMAELMAEAVVKVHMNGNWTRNTNIYPTASADNRGSILTINH 600
Db	576 IDRAEQITASVNLSESKMAELMAEAVVKVHMNGNWTRNTNIYPTASADNRGSILTINH 635
Qy	601 EAGINSLLEINGDEKVKVSGQIKSKPSDQTKWERDVEARKEPEQCVPTLVGY 660
Db	636 EAGINSLLEINGDEKVKVSGQIKSKPSDQTKWERDVEARKEPEQCVPTLVGY 695
Qy	661 DPEGSTLSSM1YPAAMYGAYGFTSDDSOILSDNDCOLQDVDTKEGQURFLANHFRANTVYN 720
Db	696 DPEGSTLSSM1YPAAMYGAYGFTSDDSOILSDNDCOLQDVDTKEGQURFLANHFRANTVYN 755
Qy	721 KFHINVPTEQSPTOATLVCNNKILDTSLTPAPEGLTYTNGQALPAKENEGCTIVSVNSG 780
Db	756 KFHINVPTEQSPTOATLVCNNKILDTSLTPAPEGLTYTNGQALPAKENEGCTIVSVNSG 815
Qy	781 KRYC1LPVCGORSYSLPDMIVGQEVYDGSAKAKVYLSDWNLYSYNRIGETVGNYNPADMK 840
Db	816 KRYC1LPVCGORSYSLPDMIVGQEVYDGSAKAKVYLSDWNLYSYNRIGETVGNYNPADMK 875
Qy	841 KIAWNGOYLDSEPRSHRVYTK 863
Db	876 KIAWNGOYLDSEPRSHRVYTK 898

KW	Hydrolase.	SEQUENCE	1002 AA;	114645 MW;	6A239DFB6408407F CRC64;				
Query Match	Score 14.9%;	DB 2:	Length 1002;						
Best Local Similarity	24.0%;	Pred. No.	1.9e-36;						
Matches	230;	Conservative	157;	Mismatches	40;				
QY	4	NSAIFYNTSQINDLOGSLAAEVKFAOSOILPAHPKGEDSQSPLHFTSLRKLILLVLP-VKA	62						
Db	137	NEIDYIYTFTEIKNVIGSLEGEVRFQTHVI-SPEGRKNPEITGRDALILFRPSIKN	194						
QY	63	DDKTPVQEARDDDNNKILGTLTLYPPSSLPLDPTYHLGDVPECGIDFTPHNGTKIKINTVA	122						
Db	195	SSSIILMIKYESDG---LTSKVNNKSPMSMLPKTDQPID-----ID-----	230						
QY	123	EYVNLDSAGSSTISHLTNALVEITANGRMVVDYLQPQDPLEGKMRVPSVAGSST	182						
Db	231	-----	230						
QY	183	VFYGDRKVTLVSYGNTLLEFKYVNGQWFRSGELENRR-ITYAHOIWSAELPAHNTVPGNLV	241						
Db	231	-----	-----	-----	-----				
QY	242	IKGNLNSGRL---NDIKIGAPGELLHTIDIGHLTTPDRDFARDKEAHR-EYFOTI	295						
QY	260	FEDEN--GNLGIIESERIKFSASELIXONIDGMLXKPRGRNIVIKELERTAVDFQKV	317						
QY	296	PYSRMTVNYNAPHLKEYWFLPQHGPGLPQHGSSTHQRIGELYHGIDRANYG	355						
Db	318	PYKLFIDESDYPHIFEXITLPNSVTEKSADIGGWQGDARBEAVGALYSTGINNANLG	377						
QY	356	LNSTAGLGENSESHPYVVAOLAAMHSRGNYANGIOWHGGGGCIVLDSTLGNESHEVGH	415						
Db	378	IYASSGY-SQ-QYVRLTNHTIAHTNIGLYNNNGTVHGSSEGGG1LVTLENTLINEWSHELGH	436						
QY	416	NYGLJHYDGFKSVRSVRAENNNTWGDGDKKRFPINPFYPSQTNNEKSCL--NNOQEPF	473						
Db	437	NYGLGHYAG--GTISH---GPITSWMGDGYKRFIANFDWRSPQSNIIRDNPQEYVKPF	490						
QY	474	TOMEPEYHTIDRAEQITASVNELESKE SMAELMAEYAVKVHMNGNWTRNIYIPTASAD	590						
Db	491	MDKTYLWPMASCGYDHONG1ISRTYLIHPYVARIQWLKGAVV--INNDYIAMD-E	546						
QY	531	547 LKIVYYVKT-----NFKVPKIKGVPPVTTGVDYDEPDFKIN-----PSQIYPPPTS--	591						
Db	591	NRSIL-----TINHEAGNYS--LFINGDEKTVSQ-----619							
QY	592	NYGNIFDLEKPRSESSLKGWQYTKDVNLYLDRVNTWHTMLVNKEEKCRFSYLSPKGKK	651						
Db	620	620 -----SFSDGFWKWERDVDTBARKEFQGVPPVTVG---Y	659						
QY	652	652 FEFGYEDLENKCTGSRSIHYLED--KKNP1--ESKYNDYF--LLSDGDEIS	701						
Db	660	660 YDPGCTL-----SSYIYPAMYGA-----YGFYSDSDQNL-----	690						
QY	702	702 YPDSTIGESK1SK1LKMSTVIGAFIGNSCQOIDGYFMNGTWAFLNQSCVSNTYW	761						
Db	691	691 DNDCOLQDVTKBGOFL-FRLANRANTYMKFHNPVTEQDQTAFLVCNNKILDTKSL	749						
QY	762	762 SNECVLKIKDNNTESISIPNTRIEKNOSKIHNLNSREKPIIDINYCGEHELTSTKV	821						
QY	750	750 TPAFPLGLTYVNGQALPAHNSGKRYCLPVGQSRGIVSLPDMVGEVYVDSG	809						
Db	822	822 SDNPD-----IKLKGPII-----VQERHGVY-----STEPKLPSG	852						
QY	810	810 AKAKVLLSDNDLNS-YNRIGEFVGNVNPADMKVKWAWNQYL---DFS-KPRSMRVV	861						
Db	853	853 ----WFKHIDNFEPEPKNE-IHELGKMRVND----NDEYICRFNFSDSDRENKKFV	897						
QY	526	526 KWNADTOEMEPYHTIDRAEQI-----TASVNELESKE	557						

Db	229	DVYTOQGD-----DIVPPFDKDAFRFLVIDAQNGQEEQEVGTTISRFPLHQAQRKAQPN	282
Qy	510	FENKAYEDRSSTSGFKNADTQEAEPEYHTIDRAEQITASVNL-----SESKNA	560
Db	283	MNNGFNLDDSSSPGTWQDQAQRYCTQVERDTDPKPOQTGVAVMTLIGIYDPYNENPSQI-	341
Qy	561	ELMAETAVVYHWN-----GWTNRNLYIPAA-----SADRQSSLTNHEAGY	604
Db	342	-----YPLVTSNQNTVNFELPOGVQAFQEGSWQPAIDLPAELESDFSWQTLMRDGEQQRY	396
Qy	605	NSYLF---INGDEKVVSYGGYKKSFSVSDGOFWKRDYDTRARKPQQFGV--PVTTLVGTY	660
Db	397	CKPTFQAANGSAVFTVGGDOS-----TRDCSSGRDLQWHINSNMTISAQGQDY	443
Qy	661	DPEGTLSSYYTAP-----YGAYGFTYSDDSQNLS-----	690
Db	444	E---LLSKYGRGAVTYTPTPEVGETLCLNKSGTDDGAGFVGNCEQISGYMHKHGK	500
Qy	691	-----DNDCQLQVDTKES--QLRFLRANHRANTVYANKFHINVPTE	729
Db	501	TWRYALRGDEVLRPSTOTOGCQLDVEFANGASERVVLNNSRHSYND--NKFHYNLAND	558
Qy	730	S-QPTQATLVCNKILDTKSLTPAPGCLTYTVNGQVLPKAENEGETVSNSGKRVCLPYG	786
Db	559	NGVPTVRLSCSDRNGETELTRFTPE-----QNPPIADLKGPPI-----IG	599
Qy	789	QRSGYS 794	
Db	600	QEXGYS 605	
RESULT 6			
Q9KTH2		PRELIMINARY:	PRT: 957 AA.
ID	Q9KTH2		
AC	Q9KTH2;		
DT	01-OCT-2000 (TREMBLrel.)	15	Created
DT	01-OCT-2000 (TREMBLrel.)	15	Last sequence update
DT	01-MAR-2003 (TREMBLrel.)	23	Last annotation update
DE	Hemolysin-related protein.		
GN	vc0930.		
OS	Vibrio cholerae.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Vibionales;		
OC	Vibrionaceae; Vibrio.		
OX	[NCBI_TAXID=666;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN/NAME / Serotype O1:		
RX	MEDLINE/200406833; Pubmed/10952201;		
RA	Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,		
RA	Dodson R.J., Haft D.H., Rickey E.K., Peterson J.D., Umayam L.A.,		
RA	Dodson S.R., Nelson K.E., Read T.D., Bettelheim H., Richardson D.,		
RA	Ermolava M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,		
RA	McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,		
RA	Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,		
RA	Fraser C.M.;		
RT	"DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae."		
RL	Nature 406:477-483 (2000).		
DR	EMBL: AE004176; AAF94092.1; -.		
DR	TIGR: JFC0930; -		
DR	TIGR; IPRO02048; EF-hand.		
DR	PROSITE: PS00018; EF_HAND; 1.		
KW	Complete proteome.		
SQ	SEQUENCE 957 AA; 103989 MW; 4C5B40FBC755CDAD CRC64;		
Query Match	6.3%; Score: 290; DB 16; Length 957;		
Best Local Similarity	53.8%; Pred. No. 2.4e-10;		
Matches	57; Conservative 14; Mismatches 35; Indels 0; Gaps 0;		
Qy	756 LTIVTNGQALPAKENEGCIVSYNSGKRYCLPYGQDPMIVQEVYDGSAKARYL	815	
Db	13 LSLSNNVVOATNEAGCJLISRINGEKYCLVGRSGTSLPSWVYHPDVQAPSWSVYM	72	



Qy	16	NDLQGSLAAEVKFAQSQILPAPKEGDSQPHLTSLRKSLLVRL---PVKAADDKTP---v	68	AC 046034; DT 01-NOV-1996 (TREMBLrel. 01, Created)
Db	131	NDMTGSTMA-----IKPDELSKG-----ITNAQDMLSKSIAKAGYVSISNDGTFCGGA	177	DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
Qy	69	QEYARDDNKNIKLGLTLYPPSSLPDTIYHLDGYPEPEGGLDFPTNGTKKIINTVAEYNK--	126	DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
Qy	178	QIRIGGSSL-----NASNDPLIVIGL-----AID---NEG.IKG.MANG.LSMVN.PAD	221	OS Clostridium difficile.
Db				Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
NCBI TaxID:1496;				NCBI TaxID:1496;
Qy	127	-----LSDASGSSSIHSHTNNALEIHTANGRWRDIYLQCPDLEGKMYFVVSSAGS	180	RN [1]
Db	222	IETLTLYKLDASATAYGRASAVCSEVIIITKGK-----NGQAPSVTYNGSYSFS	270	RP SEQUENCE FROM N.A.
Qy	181	STVFYGDKEKVTLVSGNTLLEFKVYNGQWERSGELENNRITYAHDHISWELPA-----HW	233	RC STRAIN=isolate 1470; Clostridium difficile cytotoxin B.;
Db	271	KT-----QRYRDVLSGDVEYRA-----YANQLWGDKLPALGQTANTDW	307	RL Mol. Microbiol. 17:313-321 (1995).
Qy	234	-----IVPGFLNLVIROG--NLSGRNLNDIKIGAPGELLHHTIDIGHMTRPRDFDAK	283	DR EMBL: Z23277; CAAB0815.1; -
Db	308	QQQIFRTAVSTDHVHSINGFKFLNLPYRV-----LGTTDDNGIVKTSNFR-RFTA	356	DR InterPro: IPR002479; CW_binding.
Qy	284	DKEAHREFO---TIPVSRMVY---NNYAPLUKE---VMPPTGEIJLTDMP---GNGG	330	DR InterPro: IPR001950; TIF_SUR.
Db	357	SYNIALPSFEDHUKFNTRAKFKNGKNRYADSRSRQYRWAIAITDTRPVSNEDPYQFVGGY	416	DR PROSITE: PS01110; SU11_1; -
Qy	331	WKSGTMQRIGKELYSHGIDNANTYGLNSTAGOE-----NSHPYVVAQLAHNRSR	381	DR SEQUENCE: 2367 AA; 269186 MN; EFP98223DAE70427F3 CRC64;
Qy	417	WQN-----INSTTGFSNPDKNYTSNPNSPNPLAALELKNDKG	454	Query Match Best Local Similarity 3.5%; Score 159.5 ; DB 2; Length 2367;
Db			Matches 185; Conservative 149; Mismatches 321; Indels 353; Gaps 49;	
Qy	382	NYANGIQVHGGGGGIVTLDSTGNEFSHEVHNIGL-----OHYDGFEGK	428	Qy 4 NSAIYFNTSOPENDLQGOSLAAEVKFQSQIQLPAHPKEGSOPHLTSRKSLLLYREVKA
Db	455	N-----SNDPVGNYWDYDXKFPHFLDLRHLASIGGEYAEGTQ	491	Qy 102 PEGGIDFTPHNGTKKIIINTVAEVNKLSDSA.GSSSIHSHTNNALEIHTANGRWRDIYLP
Qy	429	SVERSAENNINSTRCWGDG---KKRFITNFYPSOTNEVSCLNNOCQEFFDGKFGFDA	484	Qy 161 DTP-----VQETEADNNK1K1-----TLTLYEPSLSDPTIYHLDGV
Db	492	TIVSPYSGENNYGWNGDVTQKQYNSNIVYQYI---KSLGAND-----FDIMV	538	Qy 150 101 -----VAEILKPMNPKGS-----TNTSDSMSFLESMN --IKSFIV-
Qy	485	GGSPFSAANRFTMTPNASSAIQFFENKAVFDSRSSTGFSK---WNADTOEMEPYHTI	541	Db 1502 DSSPSEFGYISRNLIKDVYVITDNVNLLGYLKDDIKISLSLTLQDEKITKINSVHLD--
Qy	539	GSEE-----OHFHRN-----GFEEOGMDYTOE---PHDA	566	Qy 162 QGPDLEGKLMVRPVSSAGY---SSTVFIG-----DRKVTLVSYGNT
Db			Qy 1597 -----NFLQSNIKEFLDANEIIISGTSIGOEEFICDENNNIQPFIXKENTLENTLYVGNR	
Qy	542	DRAEQITASVNELESKHA---ELMAEYAVVYKVMANGNWTRNINIVPTASADNRGSLT	597	Qy 158 198 L-FKYYVQGONFRSGELEINRTRTAQ-HIWSAELPAWTIVPOLNLYTKQNIUSG-----
Qy	567	KLREQTAVATRNLVSYFGRNLNSLNNYLFETMRWOGS-----SRFSKDNRW---	615	Db 1654 QMIVIVEPNYDLDGSDISSTVINSQSKYLYGIDSCVVKVVISPNITYDEINTPTVETNN
Db			Qy 250 1713 -----RUNDIKI-----GAPGELLLHTIDIGMLTTPRDRF-DEAK	
Qy	598	INHEAGZNSYLFINGDKVSYQKKSFSVPSIDGKTFVQKWRDVKDREARKPEOFVPTTLV	657	Qy 316 1714 -----PIGEVILDANTINEKINNNINDLSIYVWSNDGNDTILMSTSEENKYSQVIRFYVNFK
Db	616	------GTFPLLAGMK---IKEENFLKDWNVLSDLKLR---LGWGTGQQ	654	Qy 284 DKE-AHRYF-----QTPVSRMIVNNYAPLHLKEYML-----
Qy	658	GYDPEGTLSXYTYPAMGAYGF---TYSDDS---ONLSDNDCCOLQDTEYREGOLFRFLANR	713	Db 1774 DKTLANKLSENFSDKQDPTVSEILSETPSYDGLJGYDGLGLSYNEKFYINNFGMM
Db			Qy 350 DNANYGLINSTAGLGENSHPYVVAQLAHHNSRGNY-----ANGIOVHGSGGGGTVL	
Qy	655	NGDDFAFAYPLVYNNNEYAOYPRDTYISTSRSRPAFNELKWEHTTNAGLDFGLNLR	714	Qy 316 1713 -----TLDEN 1918
Qy	714	ANNTYMKFHINVPTEQPTQATLVCNNKILDIKSLTPAPEGLTY-----TVNGQ	763	Db 1833 VSGLIYIIDSLYFKPPVNLITGFTVYDDKYYFNPINGGAS-----
Db	715	ITGGSDGYF-----RKTMTCTVALRSPNNTILOCDPDTEYREGKLYRGFSLNK	763	Qy 350 DNTYQGLINSTAGLGENSHPYVVAQLAHHNSRGNY-----ANGIOVHGSGGGGTVL
Qy	764	ALPARE-----NEGCTIVSVNSG---KRYCLPGV-----788	784	Qy 1883 DDKNYYFVQSG-----VLQTYGVFSTEDEGFKYFAPAN-----
Db			Db 1833 DDKKKRKRIPNYPQSO 458	
Qy	789	-----QRSGYSLPDMWVGOEYVYDGSAKARYVLLSDWNLNSYRTEFVGNVN	835	Qy 405 LGNE---FSHEYGHNTYGLHYDGFCSVRSAAENNNSTWG-----IGETI---I 1882
Db	824	FYVSQRGNNTXVQAHRVYGAANSFYVQQYDENGKPI-----ENMFVDRGN-----GTF	874	Db 1919 LEGEAIDTGTKLIDENIYVFEDNYGRAVE-----WKELOGEHMYSPE
Qy	836	PADMKKVKAUNG-----QVLDFSKPRSMR	859	Qy 459 TNEKSLCUNQCOEPFDGKFGFDAMAGGSPFAANRP-----TMYTPNSAAILORFF
Db	875	SGDKYIYKPKAGDVMLGLTSKMOYKNEFDSFSLR	908	Db 1963 -----EN 512
Qy			Db 1963 -----TGKAFKGLNQIJDXXYFNSDGFVMSORGFVSENDN 1996	
RESULT 9				
Q46034				
ID	Q46034	PRELIMINARY;	PRT: 2367 AA.	

Qy	166	--LEGAIVPRFVSSAGYSSTVFYGDRTVTLUSGVNTLFLKVNGOMFRSGLENNRITYAQ	222
Qy	1997	KHFFDSDGVMKVGTYEIDCKHFYPAENGEMIQGYFNTEDFKYFAHHNEDLGNEEGEEIS	2056
Qy	565	EYAVVKVHMNGNMTRNIYIPTASADNRGSLTINHEAGNSYLFDINGEVSYQSYKKS	624
Qy	2057	YSGYL-----NPNKNIY---FDDSFIAVGMRDLSKTYF---DEDTRAYIGLS	2103
Qy	625	FVSDQFWKFERDVDTREARKPEQFG-VPVTTLVGYDPPTGLTSSLXYIYPAAMYGAYGFTYS	683
Qy	2104	LINGQQYFNNDG1-----HQVGFYTINDKYFYFSDGILES-----GYNQI	2145
Qy	684	DDSQLNDSNDNCOLGY---DTKEGQFLRLANHRANNTVANKFHNVNPTEPSOPTATLVCN	740
Qy	2146	DDNFYIIDDONGIVQIGVPDFSDGKXYPAPA----NTVNDIY-----	2183
Qy	741	NKLDTKSLTPAPE-----GLTYYNGOALPAKENEGCTVSNSKGSKRYCLPGVGRS-GY	793
Qy	2184	QDAVEYSLVRGEDEVYVFETYETGWWYDMENE-----SDKYYFVPETKRACKGL	2236
Qy	794	SPLDPIVQGTYVDSGAKAKVLLSDADNLTS-NRIGEF-VGYNVPAD	838
Qy	2237	NLDDI---KYYFDEKGINTGLISFENNNYFENGEIOFGYINIED	2281
RESULT 10			
D	OBIL70	PRELIMINARY;	PRT: 2204 AA.
C	OBIL70:	01-MAR-2003 (TREMBREL 23, Created)	
T	T	01-MAR-2003 (TREMBREL 23, Last sequence update)	
T	T	01-MAR-2003 (TREMBREL 23, Last annotation update)	
E		Hypothetical protein.	
PF14_0379.		Plasmodium falciparum (isolate 3D7).	
S		Eukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
X	X	NCBI_TaxID:36329;	
X	X	MEDLINE:22255705; PubMed:12368864;	
X	X	SEQUENCE FROM N.A. STRAIN=3D7;	
X	X	STRAIN=3D7;	
Gardiner M.J., Hall N., Fung E., Berriman M., Hyman R.W., Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K., Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S., Chan M.-S., Nenn V., Shallom S.J., Suh B., Peterson J., Peterman J., Perceba M., Allen J., Selegut J., Haft D., Mather M.W., Valdya A.B., Martin D.M., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A., McFadden G.I., Cummings J.M., Subrahmanyam G.M., Mungall C., Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W., Fraser C.M., Barrell B.; "Genome sequence of the human malaria parasite Plasmodium falciparum." Nature 419:498-511(2002). EMBL: AEU04822; RAN6992; 1.			
R	R	Hypothetical protein.	
W	W	SEQUENCE 2204 AA; 259032 MW; 1A7E3B953FA503ED CRC64+;	
Q	Q	Query Match Similarity 3.4%; Score 158.5; DB 5; Length 2204; Best Local Similarity 18.6%; Pred. No. 0.42; Gaps 48; Matches 184; Conservative 139; Mismatches 359; Indels 307; Gaps 48;	
RESULT 11			
ID	Q97F62	PRELIMINARY;	PRT: 1157 AA.
AC	Q97F62;		
DT	01-OCT-2001 (TREMBREL 18, Created)		
DT	01-OCT-2001 (TREMBREL 18, Last sequence update)		
DT	01-MAR-2003 (TREMBREL 22, Last annotation update)		
DE	Fusion of alpha-glucosidase (Family 31 glycozyme) and Glycosidase (Treh/Mals family).		
GN	CAC281.		
OS	Clostridium acetobutylicum.		
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;		
OC	Clostridium.		
OX	NCBI_TaxID:1488;		
RP	SEQUENCE FROM N.A.		

STRAIN-ATCC-824 / DSM 792 / VRM B-1787;	Db	681	NYGEKIKITDVYDGFPGNEASSFDFYDDGTSKNEYFQDKMTLERAQDLKSVQFN	740
MEDELINE-21359325; PubMed-11465286;	Qy	651	VPTVTLIGYDPESTLSSIXYP-----AMYGAYGFTYSDDSONLSNDNC	694
Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I., Daly M.J., Tatusov R.L., Sabath F., Doucette-Stamm L., Soucaulte P., Daly M.J., Bennett G.N., Koonin E.V., Smith D.R.; "Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum.," J. Bacteriol. 183:4823-4838 (2001). EMBL: AE007786; AAU: AKA80833.;_;	Db	741	ISPKT-GYY-KSDKRNLYTKMHYKSSGDVTVGRRITRYASY-----DELKNQQEGY	791
InterPro: IPR0003122; Glyco_25.	Qy	695	QLQVDKXEGQFLRFL ANHRAANNTYPAFKFHINVPTEQPTQATLVNNKL-----DTKSL	749
InterPro: IPR00051; TonB_BoxC.	Db	792	VVGTDTGGSVYYIKVSAGHDKN-----INV-----INV-----INV-----INV	830
Pfam: PF01055; Glyco_hydro_31; 1.	Qy	750	TPAPEGLTYYTQALPAKEN-----EGLIVSNSGSKRYCLPVGORSGYSPDWI	799
PROSITE: PS00430; TonB_DEPENDENT_REC_1; 1.	Db	831	GGTYT SPQKSLKASDPNAALYTYLDGTTARTVNSTK-----TGP-----TGP	870
RW Hydrolase; Complete proteome.	Qy	800	VQQEYVYDSGAKAKVLLSDMDNLSYTRIGEFVGNYNPADMKVKRAWNGOYLDFSKPRSMR	859
SEQUENCE: 1157 AA; 129411 MW; 10B1728D833F7578 CRC64;	Db	871	-----ANGESDV-----FTEQITY-----IK 903	903
Query Match Score 3.48; Best Local Similarity 19.3%; Pred. No. 0.21; Matches 198; Conservative 120; Mismatches 335; Indels 371; Gaps 56; Sequence 1157 AA; 129411 MW; 10B1728D833F7578 CRC64;	Qy	860	VYTK 863	863
Query Match Score 156.5; DB 16; Length 1157;	Db	904	VHYK 907	907
RESULT 12	Q9F931	PRELIMINARY;	PTT; 2367 AA.	
VRADDKTPVQEARDNNK-----ILGTLTLYPPSSLPDYIYHLDGVPEGGIDFTPHNGTK 115	TD	Q9F931;		
VRADTNKVANKASKSSOKFHAKLNGNTLKKGRDETIIRICPOVFKDYKP NGKS 93	AC	Q9F931;		
KLINTVAEVNLKSLDASGSSSIHSLTNNALEIHTANGRWYD-----YLPOGSPD 166	DT	01-MAR-2001 (TREMBLrel)	16. Created	
SKDTLVDPNK-KWSTGIVSSDIKSDMV-----ITAKKMKVLAINKEDLSILVY-----DL 144	DT	01-MAR-2001 (TREMBLrel)	16. Last sequence update	
QGKLLKODSTARKTASETHNSDR-----FYGINGYNFQEDDSKGMLNRNGTES 193	DT	01-MAR-2003 (TREMBLrel)	23. Last annotation update	
TYAQHISWAELPAHWIVPGNLNVKQGNLSGRLND-----IKIGAGEI 262	TCDB			
VYAGYQGHGCSPPVWSNDGYLUYDSDGGSFTLGDTSLKYGDSKTKDTDYYVMGNKEI 253	RN			
LLHTDIG----MLTPPDRD-----DEAKDKEAHREY -FOTIPYSRM 301	RC			
LSEEDSVSKGKAPKFPKWRANGFTNTQWGMDSLSTGTGDEAKLKVINTYRSKOLPIDNFC 313	RA			
V-----NNYAPLHLKEVMLP---TGELLTDMDPGNGWHSGMQRRI-----GKE 343	RN			
LDDDWKKGQDNNGEFTKWTNDNPDSQNGQLRAYMD-SKGKRTMGIMPRILADSKGRY 372	RN			
LVSHG-----IDNANYGGLNSTAGLGENSHPKVVAQLAHHNSRGNTANG 386	RN			
VTEKGWNLPGDSEADYCGSKAKMENVPAL-----POY-----RKWANN 412	RN			
IQVHGSGGGGIVLTDSTLGNESHEVHNYGLGHY-----VDFGFKGSYHRSAAENN 437	DR			
IQ-----GAFKIGIVF-----WNDECENVNGNFGMSGD-----ISTGFDSMANQERMLSAVN-----LGFAK 506	DR			
FGF-----AMAGGSP-----FSAAANRFTMTPNSALLQREF-----ENKA-----VFDOSRSTGCF 525	DR			
WGDGKKRIPPNYPSOPENK-----SCUNNOCCBPFDHK 477	DR			
QRTWLSLNRRNYYAGAQRTSYGMMSGD-----ISTGFDSMANQERMLSAVN-----LGFAK 506	DR			
KGFD-----AMAGGSP-----FSAAANRFTMTPNSALLQREF-----ENKA-----VFDOSRSTGCF 525	DR			
WGDGKKRIPPNYPSOPENK-----SCUNNOCCBPFDHK 477	DR			
KWNADTOQEMEPYHTIDREQITA-----SYNELSESKEMLMAEAYAVKYM 573	DR			
WNGWTRNRYIP-----TASADNRGSILTINHEAGNSY-----LFING-----612	DR			
KAMOLRTYUJLPSYDRSASOSGLYRSLAMEYPNDSNAANDKEAMFGDYLMSPEV 621	DR			
QEGQTSKSKYLPPEGNWIDYTGREYTG-QTINYAVDSTNWSDIPLEFIKSGALIPTQDFE 680	DR			
WSDQFWKERDVTREARKPQEF 650	Qy	198	L-LFKYKNGOWFSGELENNTYQ-HWMSAELPAHMIVPGLNLVIKGQNL-SG-----249	249

Db	1654 QNMIVEPNYDLDSDDISSTVINFQSKYLGYIDSCVNKVVISPNIYTDEINITPVYETNN	1713	DR InterPro; IPR001254; Ser_protease_TRY.
Qy	250 -----RNDIKI-----GARFELLHITIDIGMLTTPDRF-DEAK 283	DR Pfmw; PF03422; CBN_6; 2.	
Db	1714 TYPEIVTDANYNEKINVINDLSTRYWSNDGNDFILMSTEENKVQVKIRFVNVEK 1773	DR Pfmw; PF01055; Glyco_hydro_31; 1.	
Db	284 DKE-AHREYF----QIPIVSMIVNNYAPHLKEVML---- 315	DR PROSITE; PS00135; TRIPSIN_SSRI; 1.	
Qy	1774 DKLTLKLSFNSDRODPVPEBIL-SETSYEGLIGDLYSLNFKYINNEGM 1832	SO SEQUENCE 1284 AA; 139128 MW; 2167B0DE84F42B9C CRC64;	
Db	316 -----PTGEILDT----MDPGNGWHSSTMORIGKELYSHGI 349	Query Match 3.2%; score 149; DB 2; Length 1284; Best Local Similarity 18.8%; Pred. No. 0.76; Gaps 42; Matches 170; Conservative 105; Mismatches 321; Indels 308; Gaps 42;	
Qy	1833 VSGLIYINDSLXYYKFPPVNNLITGFVTVGDDKYYFPNPINGAAS---IGETI----I 1882	Qy 82 TLTLXPPSSLPDTIYLHLDGYPEG -GIDPTEPHNTCKKINTVAEVNLKSDAGSSHTSH- 138	
Db	350 DNANYGINSTAGLGENSHPYVACLAHHNSIGNY----ANGIQVHGSSGGGIVTLDST 404	Db 54 TLVDNGAEPSSDPLLIVQAVONGILKVDIPRNSTPSAKTPMLOPNKTSVAGATINTA 113	
Qy	1883 DDKNYTQNQS-----VLQTGVESTDGFEXYFAPAN-----TLDEN 1918	Qy 139 -----LTNNALVEI-----HTANGRNWRYDLPLGPDPDLEGKMYFVSSAGYSSSTVF 184	
Db	405 LGNE--FSHEVGHNYGLGHVDPGKSVHRSANNSTKWC--DGDKKRKFIPNFYPSQ 458	Db 114 NPMPTTTSNMKIEITKPNVPRMTVKADGTL - FMEPSGGGVFSDGVFLHATGDN--M 168	
Qy	351 KAYFDSRS--STGFSK-----WNADQEMPFYEHITDRAEQITASVNESEKMAELMA 564	Qy 185 YGDRKVTLSVGNTLFLFKYNGQWPRSGELENRNRITYAOHIIWSAELPAKAVIPGLNLVTKQ 244	
Db	1919 LEGEAIDTFGKLIIDENIYEDNTYRGAVE-----WKELDGENHYSFSE-----	Db 169 YGIRSF-----NAFDGGDLIRNSHNHAH-----	
Qy	459 TNEKSLINNOCCOEPFDGHKFGEAMAGGSPIPSAANRF--TMYTPNSAITORFF---EN 512	Qy 245 GNLSGRNLNDIKIGAPGELLHITDGMALTPRDRFDFAKDEAHREYFQ--TIPVSRMIV 302	
Db	1963 -----TGAKAKGLNQIGDDYYFNSDGVCMKGFSINDN 1996	Db 194 --AGEQQD---SGGPILWSTAGYGLVDSDGKYPDTSTGOMEFYGGTPPEGRYA 246	
Qy	2057 YSCL-----WNADQEMPFYEHITDRAEQITASVNESEKMAELMA 564	Qy 303 NYAPIHL-----KEYMLPGEILITMDMPGNGNHSGTAR-----Q 338	
Db	1997 KHFDDSGVMKYGTYEDGKHYFAENGEMQIGVENTEDGKYFAHNEDIGNEEGEIS 2056	Db 247 KQNVYEYIMUGTPKEIMTDYGE-TLGKPKMLPKWLSLGFANFEDWNOTZPTNNVDTYAK 305	
Qy	565 EYAVVKYHMMNGNWTRNLYITPASADNRGSILTINHEAGYNSYLFTINGDEKVVSGYKKs 624	Qy 339 RIGKELVSHGIDNANYGLNSTAGLGENSHPYVVAQLAHNSRGNYANGLOVHG-----391	
Db	2057 YSCL-----WNADQEMPFYEHITDRAEQITASVNESEKMAELMA 564	Db 306 NIPIDAYAFDWMKYGETYGEFAWNNTNFPSSASTTSLKSTD- AKGKIMKGITKPRIV 364	
Qy	625 FVSDQFQFWKEDVVDTREARKPQFGQ-YVPVTLVGYDPEGTLSKSIYPAAMYGAGFTY 683	Qy 392 -GSGGGGIVT--LDSTLGNESFREVHNGLGHYDGFKGSVHRSAE--NNNSTWGNDG 445	
Db	2104 LINDGOIYFNDDCI-----MQVGVFTINDKVFIFSDSGTIES-----GVQNI 2145	Db 365 TKDASANVTOFGTDATNGGYF--YPOHN---EYODFIFPVTVSIDIYDYNANERAWFTH 418	
Qy	684 DDSNLSONDCQLQV---DTKEGQLRFLRANHRANNTYMKHINVPTESOPTQATLVCN 740	Qy 446 DK---KRFIPFYPSCONEKS-----CLANOCOPGPDKIKFGEPMAGGSP-- 488	
Db	2146 DDNFYIDDNGIVYQIGFDTSDGKYKAPA-----NTVNDNLY-----2183	Db 419 STDALNKIGIVGWNNDETKVSQGGALWFGNFTGHSMTQMYEGR---AYTSGAQRTW 474	
Qy	741 NKYLDTKSLTAAPE----GLTYTVNGOALPAKENEGCIVSYNSGKRYCLPGQRS-GY 793	Qy 489 -----FSAANRF--TMYTTP-----NSSAIQORFFENKAFFDSRSSTGFSKNN 528	
Db	2184 GOAVEYSGLVRVSEDVYFGETTIVTGWYIDRENE-----SDKYYFDPPTKACKGI 2236	Db 475 QTARTFYPGACYRATLWSDIGIOYNGKERINWAGHQ---EQAVMLSVNNQVVKNG 531	
Qy	794 SLDWIVCQETVTDGAKAKVLLSDWONLRY--NRIGEF-WGMVNPAD 838	Qy 545 EQI-----TASVNELESSEKMAELMAY--AVVK--VHAWN-GNW-----578	
Db	2237 NLIDDI---KYYFDEKGIMRTGLISFENNNTYFNENGEMOPGYINTED 2281	Db 592 IQRLSULIPYKAYERSAYENGGLYVPLMOAYPTDAVKNYTDAMFGDWLILLAAPVVDK 651	
Qy	RESULTS 13	Qy 579 ---TRNYYTIPASADN--RGSTLT-----INHEAGYNSYLFLINGDERKVSGYKSF 625	
Q8RQ09	PRELIMINARY;	Qy 652 QOTSKDYLPSGSWIDYARGNATGCTIRTYSVNPDLTDPFLI-----KGA 700	
ID	PPT; 1284 AA.	Db 652 VSDGQFWKEDVVDTREARKPQFGVYPTLVG--YDPEPEGLSSLYIYPAAMYGAGFTY 683	
AC	OBIOU9;	Qy 701 LIPTQ-KVQDYVGOASVTSYDVFPTDQTQSSFTYDDG-----ASYNYESGTYF 750	
DT	01-JUN-2002 (TREMBLeL 21, Created)	Db 684 DDSQNLSDNDCOLQVDTKEQJQURFLRANNTYMKHINVPTEQATLVCNK1 743	
DT	01-JUN-2002 (TREMBLeL 21, Last sequence update)	Db 751 KONMATQDNC-----SGSISFTLQAKSCTSYPALQSY1-----VKLHGSAGSTVNN- 798	
DE	6-glucosyltransferase.	Qy 744 LDTKSLTPAPEGLTYTYNGOALPAKENEGCIVSVNSGKRYCLPVQSGYSLPDWIVQGE 803	
GN	CRS2.	Db 799 -----AAITSTASLEAKAARGEG-----WATGKD 823	
OS	Bacillus globisporus.	RN [1] SEQUENCE FROM N.A.	
OC	Bacteria; Firmicutes; Bacillales; Planococcaceae; Sporosarcina.	RC STRAIN=C1;	
OX	NCB_RaxID_1159;	RA Maruta K;	
RN	"Cloning and sequencing of the genes encoding cyclic tetrasaccharide-synthesizing enzymes from <i>Bacillus globisporus</i> C11."	RT Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.	
RL	AB073929; BAB88404.1;	DR InterPro; IPR005084; CBN_6.	
DR	InterPro; IPR003322; Glyco_hydro_31.	DR InterPro; IPR003322; Glyco_hydro_31.	





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 8, 2003, 16:20:57 ; Search time 21.7788 Seconds  
(without alignments)  
3912.317 Million cell updates/sec

Title: US-10-002-309B-2  
Perfect score: 4709  
Sequence: 1 MKLKLSTCLTAPLAIGVFS.....AWNGQYLDFSKPRSMRVTIK 886

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post processing: Minimum Match 0%  
Maximum Match 100%

#### Listing first 45 summaries

Database : PIR76:\*\*

- 1: Pir1:\*\*
- 2: Pir2:\*\*
- 3: Pir3:\*\*
- 4: Pir4:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query	Match	Length	DB ID	Description
1	4709	100.0	898	2	T2131	probable toxR-regulated lip
2	2604	55.3	587	2	T00316	toxR-regulated lip
3	689.5	14.6	1002	2	T09438	toxR-activated lip
4	689.5	14.6	1013	2	B82276	ToxR-activated gen
5	671	14.2	1335	2	A82494	hemolysin-related
6	290	6.2	957	2	H82261	alpha glucosidase
7	166.5	3.5	2117	2	F91177	alpha glucosidase
8	161	3.4	1038	2	JC027	115K outer membrane
9	159.5	3.4	2367	2	S70172	toxin B - Clostrid
10	156.5	3.3	1157	2	F91255	fusion of alpha 91
11	154	3.3	1211	2	S0317	alpha, alpha-trehal
12	153.5	3.3	2366	2	S0317	toxin B - Clostrid
13	148.5	3.2	2468	2	A82412	hypothetical prote
14	148	3.1	1021	2	I40805	collagenase - Clo
15	147.5	3.1	1530	2	AH1396	peptidoglycan anch
16	147	3.1	1377	2	D90538	hypothetical prote
17	147	3.1	1366	2	AC1533	surface protein (L
18	138.5	2.9	973	2	B86547	polymorphic outer
19	138.5	2.9	973	2	F72076	polymorphic outer
20	138.5	2.9	995	2	CB1593	polymorphic membr
21	138.5	2.9	1268	2	B99189	hemagglutinin/hemo
22	138.5	2.9	1270	2	E85649	hypothetical prote
23	138	2.9	2894	2	C6474	hypothetical prote
24	137.5	2.9	2529	2	B64635	toxin-like outer m
25	137.5	2.9	3705	2	AD0123	probable autotrans
26	136	2.9	2167	2	AF1489	cell wall-associat
27	135	2.9	1579	2	B91290	probable invasin l
28	135	2.9	1700	2	G86131	probable invasin 2
29	135	2.9	2051	2	S34688	enoyl-acyl-carrie

#### ALIGNMENTS

##### RESULT 1

T42131

Probable toxR-regulated lipoprotein tagA - Escherichia coli Plasmid pO157

C:Species: Escherichia coli

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Nov-2000

C:Accession: T42131; T00210

R:Burland, V.; Shao, Y.; Perna, N.T.; Plunkett, G.; Sofia, H.J.; Blattner, F.R.

Nucleic Acids Res, 26, 4196-4204, 1998

A:Title: The complete DNA sequence and analysis of the large virulence plasmid of Esc

S.: Makino, K.; Ishii, K.; Yasunaga, T.; Hattori, M.; Yokoyama, K.; Yatsudo, H.C.; Kubo

DNA Res., 5, 1-9, 1998

A:Cross-references: EMBL:AF074613; PIDN: AAC70099.1

A:Experimental source: strain EDL333; serotype O157:H7

R: Makino, K.; Ishii, K.; Yasunaga, T.; Hattori, M.; Yokoyama, K.; Yatsudo, H.C.; Kubo

A:Reference number: Z22068; PMID: 98391744; PMID: 9722640

A:Accession: T42131

A:Status: preliminary; translated from GB/EMBL/DDJB

A: Molecule type: DNA

A: Residues: 1-898 <BUR>

A: Cross-references: EMBL:AB011549; NID: d1204561; PIDN: BAA31757.1; PID: d1032718

A: Accession number: 214127; MUID: 98290340; PMID: 9628576

A: Status: preliminary; translated from GB/EMBL/DDJB

A: Molecule type: DNA

A: Cross-references: 569-898 <MAK>

A: Accession number: 214120

A: Status: preliminary; translated from GB/EMBL/DDJB

A: Molecule type: DNA

A: Cross-references: 886-886 <MAK>

A: Accession number: 214127; MUID: 98290340; PMID: 9628576

A: Status: preliminary; translated from GB/EMBL/DDJB

A: Molecule type: DNA

A: Cross-references: 886-886 <MAK>

A: Accession number: 214127; MUID: 98290340; PMID: 9628576

A: Status: preliminary; translated from GB/EMBL/DDJB

A: Molecule type: DNA

A: Cross-references: 886-886 <MAK>

A: Accession number: 214127; MUID: 98290340; PMID: 9628576

A: Status: preliminary; translated from GB/EMBL/DDJB

A: Molecule type: DNA

A: Cross-references: 886-886 <MAK>

A: Accession number: 214127; MUID: 98290340; PMID: 9628576

A: Status: preliminary; translated from GB/EMBL/DDJB

A: Molecule type: DNA

A: Cross-references: 886-886 <MAK>

A: Accession number: 214127; MUID: 98290340; PMID: 9628576

A: Status: preliminary; translated from GB/EMBL/DDJB

A: Molecule type: DNA

A: Cross-references: 886-886 <MAK>

A: Accession number: 214127; MUID: 98290340; PMID: 9628576

A: Status: preliminary; translated from GB/EMBL/DDJB

A: Molecule type: DNA

A: Cross-references: 886-886 <MAK>

A: Accession number: 214127; MUID: 98290340; PMID: 9628576

A: Status: preliminary; translated from GB/EMBL/DDJB



				A; Status: preliminary
				A; Molecule type: DNA
				R; Residues: 1-1013 <HEI>
				A; Cross-references: GB:AB004167; GB:AE003852; NID:99655268; PID:AAF33983.1; GSPDB:G
				A; Experimental source: serogroup O1; strain Ni6961; biotype El Tor
				C; Genetics:
				A; Gene: VC0820
				A; Map position: 1
				Query Match 14.6%; Score 689.5; DB 2; Length 1013;
				Best Local Similarity 24.2%; Pred. No. 2.1e-36;
				Matches 232; Conservative 151; Mismatches 279; Indels 295; Caps 39;
			OY	27 NSAIYFTSQPNLDSLAAEVKFAOSQILPAHPKEGDSOPHLSLRKSLLLVRP-VKA 85
			Db	148 NEIDYHTTEIEKVNIGSLEGEVRYQTHVI--SPEGRKNEPEITIGRDALILEFSPSKN 205
			QY	86 DDKTPYCOEARDDNKKNLGLTLYPPSSLPDITYHLDGVPEGGIDEFPHNGTKKINTVA 145
			QY	206 SSSILMKIYSED--LTSKRVPAKSKPMLPKTDQPID----ID----- 241
			Db	146 EVNKLSASGSSTIHSHTNNALVEIHTANGRWRDYLQPQSPDLESKMRVTTSSAGY SST 205
			Db	242 ----- 241
			QY	206 VFYGDRRTVLSVGNTLFLKYKYNQWTFSGELENRR-ITYAHOIWAELPAWIVPOLNLV 264
			Db	242 ----- ENNKVSYNSNSYWSAELPWNMKKSMLSH 270
			QY	265 IKOGNLSSRL---NDIKIGAPGELLHTIDIGMATTPRDRFDANDKEAHR-EYFQT 318
			Db	349 PYKLISSDYTPHFETITLPNTVTEKSAADIGGWHDREAVGIALVSTGINNALG 388
			QY	271 FEDEN--GNLGIESERIKFSAPSSELIONIDGLMLWKPRGRNIVIKELETAVDVFQKV 328
			Db	319 PYERMVINYNAPHLKEYMLPTGELLTMDPONGNHSGTHORIGELYSHGIDIANYNG 378
			QY	389 IVASSGTSQ-QFNRLTWHITAHTNIGYNNGYVHGGSGGGIVTLNTLINEWSIELGH 447
			Db	502 MDKTYLTDAMSGYDONGIIISRYTLHPPYARIODWLKGAVV--INNDYMWDF-E 557
			QY	439 NYGLGHYDPEKTFPNTVTEKSAADIGGWHDREAVGIALVSTGINNALG 388
			Db	379 LNSTAGLGENSHSPVVAQLAHHNSRGNYANGLOVHGGSGGGIVTLDTLIGNEFSEVGH 438
			QY	547 LKNIYVYKT-----NFKVPKKGKPYVVTILGYDDEDKIN-----PSQLYPPTIS-- 591
			Db	389 IVASSGTSQ-QFNRLTWHITAHTNIGYNNGYVHGGSGGGIVTLNTLINEWSIELGH 447
			QY	614 NRGSILTIN-----HEAGYN-----SYLFING-- 635
			Db	502 MDKTYLTDAMSGYDONGIIISRYTLHPPYARIODWLKGAVV--INNDYMWDF-E 557
			QY	554 TQEMEPYEHITDRAQTAVSNELSESKMAELMAEYAVKVMNGNTRNIYIPTASAD 613
			Db	389 IVASSGTSQ-QFNRLTWHITAHTNIGYNNGYVHGGSGGGIVTLNTLINEWSIELGH 447
			Db	502 MDKTYLTDAMSGYDONGIIISRYTLHPPYARIODWLKGAVV--INNDYMWDF-E 557
			QY	592 NYGNIFDLEKPRSESSLKQWQYYKDVNLYLDRVNTHWHTMLVNREEKICRFSYLSPKGKK 651
			Db	389 IVASSGTSQ-QFNRLTWHITAHTNIGYNNGYVHGGSGGGIVTLNTLINEWSIELGH 447
			QY	636 -----DEKVTSGQGKKSFSYDQFMKMERDVDTATEKRYPEQFGPYTLYG-----Y 682
			Db	502 MDKTYLTDAMSGYDONGIIISRYTLHPPYARIODWLKGAVV--INNDYMWDF-E 557
			QY	652 FEFLGYEDIENKIKTGGRSIHYLEDG--KKNP1----ESKNDYF--LLSDGDGEIS 701
			Db	389 IVASSGTSQ-QFNRLTWHITAHTNIGYNNGYVHGGSGGGIVTLNTLINEWSIELGH 447
			QY	683 YDPGCTL-----SSYIYPAMYGA-----YGFYSDDSQNLIS-- 713
			Db	502 MDKTYLTDAMSGYDONGIIISRYTLHPPYARIODWLKGAVV--INNDYMWDF-E 557
			QY	702 YVPDGTSKICSLKMSCTVYGAFFIGNSCROIDGYFMNGQWAFTLNQSSVNTYTW 761
			Db	389 IVASSGTSQ-QFNRLTWHITAHTNIGYNNGYVHGGSGGGIVTLNTLINEWSIELGH 447
			QY	714 DNDCQLQVDTKEGOLR-FRLAHNHNANTVNKHINVTESQPTQATLVNNILDTKSL 772
			Db	502 MDKTYLTDAMSGYDONGIIISRYTLHPPYARIODWLKGAVV--INNDYMWDF-E 557
			QY	762 SNECVLKIKDDNNEIESIPNRYIEKNOSKHNISREKPIIDINVYCGEHELTISKV 821
			Db	389 IVASSGTSQ-QFNRLTWHITAHTNIGYNNGYVHGGSGGGIVTLNTLINEWSIELGH 447
			QY	773 TPAPPGLTVNVQALPAKENEGCTVSNSGKRYCLPQQRSGSYSLPDWVQGEVYDSG 632
			Db	502 MDKTYLTDAMSGYDONGIIISRYTLHPPYARIODWLKGAVV--INNDYMWDF-E 557
			QY	822 SDNP-----IRKLKGPII-----VQGERGY-----SVEPKLPSG 852
			Db	389 IVASSGTSQ-QFNRLTWHITAHTNIGYNNGYVHGGSGGGIVTLNTLINEWSIELGH 447
			QY	833 AKARYULLSDMDNLS-YNRIGEFYGNPADMINKVKAWNGQYL---DFS-KPRSMRVMV 884
			Db	502 MDKTYLTDAMSGYDONGIIISRYTLHPPYARIODWLKGAVV--INNDYMWDF-E 557
			QY	853 -----WFKXIDNFPEKNE1NHELGKMRVND-----NDEYICRFNFSDSDREMKFV 897
			Db	389 IVASSGTSQ-QFNRLTWHITAHTNIGYNNGYVHGGSGGGIVTLNTLINEWSIELGH 447
			QY	636 -----DEKVTSGQGKKSFSYDQFMKMERDVDTATEKRYPEQFGPYTLYG-----Y 682
			Db	502 MDKTYLTDAMSGYDONGIIISRYTLHPPYARIODWLKGAVV--INNDYMWDF-E 557
			QY	663 FEFGLGYEDIENKIKTGGRSIHYLEDG--KKNP1----ESKNDYF--LUSIDGDEIS 712
			QY	683 YDPGCTL-----SSYIYPAMYGA-----YGTYSDDSONLS-- 713
			Db	713 YYPDSTIGESKISCSLKMSGTYSVAGFKIGNSCROIDGYFMNGFOWAFTLNGSGVNSTYT 772
			QY	714 DNDCQLQVDTKEGOLR-FRLAHNHNANTVNKHINVTESQPTQATLVNNILDTKSL 772
			Db	773 SNBCVLIKDDNNEIESIPNRYIEKNOSKHNISREKPIIDINVYCGEHELTISKV 832
			QY	773 TPAEGGIFTYTVNQALPAKENEGCTVSNSGKRYCLPQQRSGSYSLPDWVQGEVYDSG 832
			Db	833 SDNP-----IRKLKGPII-----VQGERGY-----SVEPKLPSG 863

## RESULT 4

B82276	ToxR-activated gene A protein VC0820 [imported] - Vibrio cholerae (strain Ni6961 serogroup O1; biotype El Tor)	YGTYSDDSONLS-- 713
C; Species: Vibrio cholerae		
C; Date: 18-Aug-2000 *sequence_revision 20-Aug-2000 *text_change 02-Feb-2001		
C; Accession: B82276		
R; Heideberg, J.-F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.; Charlson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.; Mekalanos, J.J.; Ventler, J.C.; Fraser, C.M.		
Nature 406, 477-483, 2010		
A; Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.		
A; Reference number: A82035; PMID:20406833; PMID:10952301		
A; Accession: B82276		

Qy	833 AKAKVLLSDPDLNLS-YNRIGEFVGNVPADMKKYKAWNGQYL---DFS-KPRSMRYY 884	817 EQGEYQLE-----GW-----QAAGD-----LT-QAEIQYQNWQTLILIDSQL 852
Db	864 -----WEKAYDNFEPKNE-NHELGKMRVND-----NUDYICRFNEDSDRNMKFV 908	639 VV-----SOGYKKSFYSDGQFWKERDVIDV-TREARKEPQFSVYPTTLYGYDPEGTLS 690
<b>RESULT 5</b>		
A82494	TagA related protein VCA0148 [imported] - <i>Vibrio cholerae</i> (strain N16961 serogroup O1)	853 PICRFDTYNTNGOSATFV--GSINAQRVCEGSRDMR-----WYNDYQIDSPPVG 899
C;Species:	<i>Vibrio cholerae</i>	Qy 691 SYIYPAMKGAAGTYSSDSQ-----NLSDND-----
C;Date:	18-Aug-2000 *sequence_revision 20-Aug-2000 *text_change 02-Feb-2001	Db 900 QYELLSOFGAVNVTYPAEIGEVOLCTLNKPENNGSDAGEFVRNGRCQEYGVKNNAE 959
C;Accession:	A82494	Qy 717 -----COLQYDTKEQQLREFLANNHRANNPHKSYHIVPTE- 752
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Dodson, R.J.; Charon, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.J., R.R.; Mukalanos, J.J.; Venter, J.C.; Fraser, C.M.	Db 960 GRYWSYARNESEVLSRTLASQRCELVHNRGNSTHALDGNRHKSYESNEKHVNLSMEK 1019	
Nature 406, 477-483, 2000	Qy 753 SOPTQATLVCNNKILDTSKSLTPAPEGLTYTNCQALPKANECCIVSYNSGRYCLPYGO 812	
A;Title: DNA Sequence of both chromosomes of the cholera pathogen <i>Vibrio cholerae</i> .	Db 1020 GVPfqVSLSCSD-LNGTSTLT-----RFPD-QNPPLKLKGPII-----IGQ 1060	
A;Reference: A82494	Qy 813 RSGYS 817	
A;Status: preliminary	Db 1061 EYGS 1065	
A;Molecule type: DNA	Qy 814 -----	
A;Residues: 1-1335 <HEI>	Db 1062 H82261	
A;Cross-references: GB:AE004356; PIDN:99657536; PIDN:AAF96061.1; GSPDB:GN001	hancholysin-related protein VC0930 [Imported] - <i>Vibrio cholerae</i> (strain N16961 serogroup O1)	
A;Experimental source: serogroup O1; strain N16961; biotype El Tor	C;Species: <i>Vibrio cholerae</i>	
C;Genetics:	C;Date: 18-Aug-2000 *sequence_revision 20-Aug-2000 *text_change 02-Feb-2001	
A;Gene: YCA0148	C;Accession: H82261	
A;Map position: 2	R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Dodson, R.J.; Chardon, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.J., R.R.; Mukalanos, J.J.; Venter, J.C.; Fraser, C.M.	
Query Match 14.2% Score 671 DB 2; Length 1335; Best Local Similarity 25.5% Pred. No. 5.2e-35; Matches 231; Conservative 10; Mismatches 248; Indels 318; Gaps 36;	Nature 406, 477-483, 2000	
Db 39 NDLOGSLAAEYKFAQSQILPAHPK--EGDSQPHTLS--RKSLLLVVRPVKAADDKTPQVQE 94	A;Title: DNA Sequence of both chromosomes of the cholera pathogen <i>Vibrio cholerae</i> .	
Qy 353 NHQGSLEGGLISITQHTSYA-PKGNELTQGHDAIMREALLFTPOQGEETINQVRAE 411	A;Accession: H82261	
Db 95 ARDDNNKILGTLTVYPPSSLDTIHLQDGYPEGDFPTNGTKXLIINTVAEVNKLSDAS 154	A;Reference number: AB2035; MUID:20406833; PMID:10932301	
Qy 412 VFLDGEVQTLMLL-PPSAL-----AASDQPENGRMKV-----443	A;Status: Preliminary	
Db 155 GSSIH3HLTNNAVLEIHTANGRWRDYIYLPGDPLEGKAYTFRVSSAGYSSTVFYGDRKVY 214	A;Cross-references: GB:AE004176; GB:AB003882; PIDN:99655388; PMID:AAF94002.1; GSPDB:GN	
Db 444 --VFSLA-----449	A;Experimental source: serogroup O1; strain N16961; biotype El Tor	
Qy 215 LSGVNTLFLFKVNGQWERSGLENNRITYAQHWSAELPAHWYIPLNLYI-----KQGN 269	C;Genetics:	
Db 450 -----W\$PLQNDWMKGLSLRITDNEIGREGV 476	A;Gene: VC0930	
Qy 270 LSGRLNDIKIGAKGEELLHTIDGMHTTPRPERFAKD-KEARHREYFOTIPVPSRMVNNY 328	A;Mapp position: 1	
Db 477 LS--QGEIOFGQAPALEVNDIQLMPPRMTIONLPTLAADYFOKPKASKLYMADY 534	Query Match 6.2%; Score 290; DB 2; Length 937;	
Qy 329 APLHKEVMTGELLTDAGDNGPHSGTSHGIDNANTGLNSTAGLGEN 388	Best Local Similarity 53.8%; Pred. No. 1.5e-20; Matches 51; Conservative 57; Mismatches 35; Indels 0; Gaps 0;	
Db 535 TPAHFPTVMTNGVYVYTDKSASTGWPBHSQMDREAIKGANYSTGINNANYGIVSSAGYSQQ 594	Qy 779 LTYTYTNQALPAKENEGCIVSYNSGRYCLPYGORSGYSLPDWVQEVYVDSGAKAKVYL 838	
Qy 389 ----SHPYVVAQLAHNRSRNYAN -----GIOYHGCGGGTIVLDSTLGNFSHEVG 437	Db 13 LSLSNVNVQQTNEAEGCTISRLNGEKYCLKVGENISGSLPSWYAHVDPQAFSGVSV 72	
Db 595 YNRRFNH-----ITAHNTVGYTPTDQLEQVYHGGSGGGIVLEATGNERSHELG 648	Qy 839 LSDNDLNSYNRIGEFVGYNPNADMKKVAMQYLDIFSKPSRSMRV 884	
Qy 438 HNYGLGHYVDGFKGSYHRSAAENNNTGWDODDKRIPNF-----YPSQTEKSCLNQ 491	Db 73 LSDNDLNSYNRLAVEDRTGNEIDLKVRYANGAYLDFSKPSRSMRVYL 118	
Db 649 RNYGLGHM--PYMASIH---DLESQWGMDFPHORETGILNHWKGDYTOQGD-----D 696	RESULT 7	
Qy 492 COEPF-DGHKFGDAGGSF--FSAAANRFTPYTPSSAIIOFFENKAVFDSRSSTGFS 548	F97177	
Db 697 TYPFDKAFRLFRDAQNGQEEYVGTISRFLEHPQSKRQWRANGFNDLSDHPSGTV 756	C;Species: Clostridium acetobutylicum	
Qy 549 KNNADTQEMPEYHTIDRAEI-----TASVNELES 580	C;Date: 14-Sep-2001 *sequence_revision 14-Sep-2001 *text_change 14-Sep-2001	
Db 757 QWDQETORYKAETDTPKPOQYGVPVVTLGLGIYDPONENPQIPIYVSYNYNFELPQP 816	R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L.J.; Daly, M.J.; Bennett, G.N.; Koobin, E.V.; Smith, D.R.	
Qy 581 KHAELMAYAVVVKHMNGNNTRNVIYIPTASADNRGSLILTNEAGYNSY -LF NGDEK 638	J; Bacteriol. 183, 4823-4838, 2001	
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium		
A;Reference number: A96300; MUID:2135925; PMID:21353325	A;Accession: F97177	
A;Status: Preliminary	A;Status: Preliminary	
A;Molecule type: DNA	A;Molecule type: DNA	
A;Residues: 1-1217 <KUR>	A;Residues: 1-1217 <KUR>	

Genetic Comparisons									
Gene: CAC2252		Cross-references: GB:NE001437; PIDN:AAK80209.1; PID:915025253; GSPDB:GN001628		Experimental source: Clostridium acetobutylicum ATCC824		Best Local Similarity 18.6%; pred. No. 0.021;		Matches 179; Conservative 121; Mismatches 319; Indels 343; Gaps 48	
Y	Y	79	LVRPVKA	--DDKTPVQVEARDNNKKLQGTLVPPSSLP	--DTIVHLDGVPEGGI	129			
Y	Y	80	KICEPQVLKVDPKP-SQSSSDTLPVNPKIWN	--TGNIISSDLSNDSEMVTTQKMTKIRIS	137				
Y	Y	130	--DETPINGTKKIINTAEVYNKLSDASGSSTTSHSLTNNALV	--EIHTA	173				
Y	Y	134	NGRWTRDYLPLQGPDELEGKMY	--RVSAGYSSSTFYGBRKVTLSVGNTILFKTYNGQW	230				
Y	Y	138	KSDLTMSVY	--DSTGKQIVKQOQIAKSXSVFTHNSGDR	181				
Y	Y	231	FR	--SGEFLN-NITYAHOIWISSLPAHRIVPGNLVIQGNLSGRLND	276				
Y	Y	182	FKEDSNGKGHRNGTTSVYRQGHGSPFWNSDQGLYDSDGGFTIGTSLQYSGS	241					
Y	Y	277	--IKIGAPGEPLLHLTDI	--GMLTTP	313				
Y	Y	242	KTDIPTYLMGSPKPEVISESDVSGKAPMFPKWAQFTNTOWGNWNSLSCGTGNDDELKLS	301					
Y	Y	314	YFQT	--TIPVSRMIV	356				
Y	Y	302	VLTNTYRSKQIPIPDNECLDFENKWKWGDNYGEFKWNTDNFDAQNGQKLKAYMD	-SKGKMT	360				
Y	Y	357	GTMRORI	--GIELVSHG	393				
Y	Y	361	GIMFRILLADSEQARYVTSGKWWLPGDSAASDYCSKGMMPNRYNA	SDV	409				
Y	Y	394	VAQLAHNSRGNYANCIQVHCGSGGGIVTLSTDLGNEFHSHEVGHVNGHLGHY	--	445				
b	b	410	--RKMWWNNIQ	--DAFDKGIVGFWDNEDECDEVNTFGNGNMNMRERA	450				
Y	Y	446	-VDGFKGSVRSAAENNNTWG	--WDGDKKRIFINFEYPSOTNEK	485				
b	b	451	IYDG	--QRAYKNQRVWSLNRRNYAGAQRYCIGHWSGD	--1STGFDMSANQRFMR	500			
Y	Y	486	-SCLNNQCQEFDGHRFGFD	--AMAGGSP	538				
b	b	501	LSAVN	--LGENAWGADTGFGNDPDTPEKARPAHAEFSAPT	--IFRVHQDNRV	550			
Y	Y	539	-FDSRSSTGFSKWNADTQEMEPYEHITDRAEQITA	--SYNELSESK	581				
b	b	551	RYPWAFGSTABAAKAMOLRVTIPIYIYSTDRSASSGGLVRLSRMKEYPNDSNAANDK	610					
Y	Y	582	MAELMREAYAVVKVHARINGNTRNITYP	--TASADNRSSILTNHEAENSY	630				
b	b	611	EAMFESDMLVSPVNRQGOTSKSIYLPEGNWIDYTGTGREYTGQ	-QTINYAVDSTNWNSDIP	669				
Y	Y	631	LFINIG	--DEKVYSGQYKKSFSVSD	658				
b	b	670	LFIKSSAIIPTQDFENVGEKIKITDVYDFAFSKATTFDYYDDGTSYDENGSKYDOK	729					
Y	Y	659	DVVDTCREARKPEQFQGYPTVYDPEGTLSYI	--Y	695				
b	b	730	MTLQTSSTDSDKSYQFNIDKNT	--GSYTPD	785				
Y	Y	696	AMYGRGFTYSDSQQLSNLDQQLVQDTCFKEQFLRFLANHRNANTVWAKFHIMPVTEQP	755					
Y	Y	786	ALKASSEGEYASGT	--DTYGNVYVYKVSQGDAKN	826				
Y	Y	756	TQATLYCNK	--ILDTSKSLTPAPEGILT	808				
b	b	827	VTITAAAPKPGSTYGPQTVS	STASKSDATIYTL	875				

809 p.v 810

**b** 809 PV 810  
876 I: 877

Qy	565 DAEQITASVNESESERMA---ELMAYEAYVKVIMWNGNTRNIYIPTASADNRGSLT 620	185 QGPDELEGAKAVPRVSSAGY--SSTVYFG-----DRKVTLSVGNT 220
Db	567 KIREQTAYATRNLTVSIFGRINYSLARRYLTFPTMRDGS---SRSKDNW--- 615	1597 ---NPLQSNUKFILDANF1ISGTSIGQEFFICDENNNIOPYRIKFNTLETTLYVGNR 1653
Qy	621 INHEAGCNSYLEINGDEKVVSGQYKKSFSVSDQGKFQKRDVYDTREARKPEQGVPVTILY 680	221 L-LFKYVNGWFERSGELENRITYAQ-HIWSAELPAHNVIPGILNVLKGQNLSG----- 272
Db	616 -----GTPSLALGWK---IKEENFLDNVNLSDLKLRL---LGWGTGQQ 654	1654 QNMTEPNDLDDSDISSTVINFQSQKTYLGDSCCNVRYVISPNVITYDEINTPVYETNN 1713
Qy	681 GYDPEGTLSSIIYPAHYGAF-TYSDDS---ONLSNDNCQLOQDTEKQGQLRFLANHR 736	273 -----RLNDIKI----GAPGEJLJLHTIDIGMLTPRDRF-DFAK 306
Db	655 NGDDDFAYPLPLVYVNNYEQAQYQFQDTYSTSRSPKAENLKEWKTTWNAGLDFGLNLR 714	QY 1714 TYPEIVLDARYINERKINNINDSIRYWSNDNFMSNKVSQV1RFVNEYK 1773
Qy	737 ANNTVANKFHINVPTEOPTQTLVCAANKILLDKSLTPAPELTY-----TVNGQ 786	QY 307 DKE-AHREYF----QTIPVSRMIVNNYAPLHLKEVML----- 338
Db	715 RTGGIDGYF-----RTKMTCTVALRSPMNLQCPDDEYKFTGKLRYGFSINAK 763	Db 1744 DKLANKLSPNSFSDQDVPSSEITL-SFPSPSYEDGLIGDGLVSYLNKEFYINNEFGMM 1832
Qy	787 ALPAKE-----NEGCVSVNSG--KRYCLPVG----- 811	Qy 339 -----PTGEELTD-----MDPGNGGWHSGTMQRORIGKBLVSHG 372
Db	764 PIVTKDFTWDLSNITYTNWHNEITKLATGDDSYVEAGDK1SQRGNTTKVQAHKVGAANS 823	Db 1833 VSGLIYINDSUYFPPVNNLITGFVTGDDKYVNPINGGAAS-----IGETI----- 1882
Qy	812 -----QRCGYSLLPDIWGVDEVYVSGAKVYLSDWLSYNTNRIEFVGNVN 858	Qy 373 DNANYGLNSTAGLGENSHPYVVAQLAHHNSRGNY----ANGIQVHGGGGGGIVTLDST 427
Db	824 FVYVSRGNNKTQVAHKVGTAAANSSTYQVQOYDINGKPI-----ENMFVDRGN--GTID 874	Db 1883 DDKNYTPNQSS-----VLTQGIVFSTDGFKYPFADAN-----TLDEN 1918
Qy	859 PADMKKKVKAANG-----QYLDSESKPRSMR 882	Qy 428 LGNE---FISHEVGHNYGLAHYVDFGFKGSTHRSAAENNNSRGW-----DGDKKRFIPNFYPSQ 481
Db	875 SGDKYIYKPKAGDYLMGLTSKMQYKNEFDSESLR 908	Db 1919 LEGEAIDFTGRLIIDIENITYFEDNRYGATE-----WKEDGENHYFSP----- 1962
Qy	S70172 toxin B - Clostridium difficile	Qy 482 TNEKSLINNOQCQEPFDHKPGFAMAGGSPFSAANRF-TMYTPRNSSAIQORFF---EN 535
C;Species:	Clostridium difficile	Db 1963 -----TGRAFKGLNQIGDQKYYNSDGTQKGTVSINDN 1996
C;Accession:	S70172; S44271	Qy 536 KAVFDSRS-STGFSK-----WNADTOQEMEPYHTIDRAEQITASVNELESKMAELMA 587
R;von Eichel-Streiber, C.; Meyer zu Heringdorf, D.; Habermann, E.; Sartingen, S.	Db 1997 KHYFDSSGVMKVGYTEIJDGHFYFAENGNOIGVNTEDGFYPAHNNDLGNEEGEITS 2056	
Mol; Microbiol 17, 313-321, 1995	Qy 588 EYAVYKHMONGNTRNIYPTASADNRGSILTINHEAGTNSYLFTINGDEKVVSGQYKLS 647	
A;Title: Closing in on the toxic domain through analysis of a variant Clostridium difficile	Db 2057 YSGIL-----NFNNKYY--FDDDSFTAVGVKMDLGSKYTF--DEDDTAEAYIGLS 2103	
A;Accession: S70172	Qy 648 FVSDGQFWKERDWDYDTREARKAPEQFCG-VPYVTLYGYDPGEGTLLSSYYIYAMYGAGFTYS 706	
A;Status: preliminary; nucleic acid sequence not shown	Db 2104 LINDGOYFNDG-----MOYGFVTFNDKTYFESDSGIES-----GVQNI 2145	
A;Molecule type: DNA	Qy 707 DDSQNLSDNDQLOQV--DTXEGOLRFRPLANHRANTVMKPHAINVPTESOPTQTLVNC 763	
A;Residues: 1-2367 <VON>	Db 2146 DDNYFYIDNG1VQGYFSDGKYYFAPA-----NTVDNVIY----- 2183	
A;Cross references: ENBL:Z23277; NID:9761713; PID:CAA80815.1; PMID:7494480	Qy 764 NKILDPKSLTAPE---GLTYTYNGQALPAKENEGCVSNSSKQYCLPGQRS--gx 816	
R;Sartingen, S.; von Eichel-Streiber, C.	Db 2184 GOAVEYSGLVRVYGEDVYYFGETYTIETGWYDME-----SDKYYFVFPETKACKGI 2236	
submitted to the EMBL Data Library, July 1993	Qy 817 SLPDWIVGQEYVVDSSAKAKYVLLSDMDNLSY--NRIGEF-VGNNYNPAD 861	
A;Description: Cloning and sequencing of an Clostridium difficile toxin B mutant.	Db 2237 NLIDDI--KYPDEKGIMRTGLISFNNNNYFENGEIOFGYINIED 2281	
A;Reference number: S44271	Qy 3 RESULT 10	
A;Molecule type: DNA	F97255	
A;Residues: 1-1323; 'N'; 1325-2367 <SAR>	C;Species: Clostridium acetobutylicum	
A;Cross references: ENBL:Z23277	C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001	
C;Keywords: cpl repeat homology	R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Matkarova, K.S.; Zeng, Q.; Gibson, R.; L	
Query Match 3.4%; Score 159.5; DB 2; Length 2367;	Daly, M.J.; Bennett, G.N.; Koontz, E.V.; Smith, D.R.	
Best Local Similarity 18.4%; Pred. No. 0.17; Matches 149; Mismatches 321; Indels 353; Gaps 49;	J;Bacteriol 183 4823-4838, 2001	
Matches 185; Conservative 149; Mismatches 321; Indels 353; Gaps 49;	A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium	
Qy 27 NSAIYFNTSOPINDLOGSLAAEYKFAQSQSILPAPHPKEGDSQPHLTSURKSLLVRYPKAD 86	A;Reference number: A96900; PMID:21359325	
Db 1454 NSELQKNIYPSFVSESEGK-----ENSPINASTKEG---LFVSELPDVLSKRYMD 1501	A;Accession: F97255	
Qy 87 DKTP-----VOVEARDDNKNLKG-----TLLTYPSSSLPDTIYHLDGV 124	A;Status: preliminary	
Db 1502 DSKEPSFGYYSNLRKDVKVTIDKVNLYGKQYLDKIKLSSLQDENTIKNSVHD-- 1559	A;Molecule type: DNA	
Qy 125 PEGG'DPThNGTKKIIVAEVNLKSLDASSSS1HSHTNNNALEIHTANGRVRDYL 184	A;Residues: 1-1157 <KUR>	
Db 1560 -ESG-----VATEILKFMRKGS---TNTSDS1MSFESMN--IKSIFY- 1596	A;Cross-references: GB:AE001437; PID:915025938; GSPDB:GN00168	



RESULT 1	
S10317	toxin B - Clostridium difficile
C; Species: Clostridium difficile	
C; Date: 30-Sep-1993 *sequence revision 30-Sep-1993 *text_change 15-Oct-1999	
R; Barroso, L.A.; Wang, S.Z.; Phelps, C.J.; Johnson, J.L.; Wilkins, T.D.	
Nucleic Acids Res. 18, 4004, 1990	
A; Title: Nucleotide sequence of Clostridium difficile toxin B gene.	
A; Reference number: S10317; MUID:9036540; PMID:2374729	
A; Accession: S10317	A; Status: translation not shown
A; Molecule type: DNA	
A; Residues: 1-2366 <BAR>	
A; Cross-references: EMBL:X53138; NID:940442; PID:CAA37298.1; PID:940443	
R; Eichel-Streiber, C.; Lauenberg-Feldmann, R.; Sartingen, S.; Schalze, J.; Sauerborn, M	A; Description: Comparative analysis of Clostridium difficile toxins A and B.
A; Reference number: S21894	A; Accession: S21894
A; Molecule type: DNA	A; Molecule type: DNA
A; Residues: 1271-2366 <ECID>	A; Residues: 1271-2366 <ECID>
A; Cross-references: EMBL:Z60984; NID:940445; PID:CAA43291.1; PID:940446	A; Cross-references: EMBL:X53138; NID:940442; PID:CAA37298.1; PID:940443
R; von Eichel-Streiber, C.; Lauenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerborn, M	Mol. Gen. Genet. 233, 26-28, 1992
A; Description: Comparative analysis of the Clostridium difficile toxins A and B.	A; Title: Comparative analysis of the Clostridium difficile toxins A and B.
A; Reference number: S22434	A; Reference number: S22434
A; Accession: S22434	A; Accession: S22434
A; Molecule type: DNA	A; Molecule type: DNA
A; Residues: 1791-2366 <VON>	A; Residues: 1791-2366 <VON>
A; Cross-references: EMBL:X60984	A; Cross-references: EMBL:X60984
C; Genetics:	C; Genetics:
A; Gene: toxB	C; Superfamily: cpx repeat homology
C; Keywords: cytotoxin	C; Keywords: cytotoxin
Query Match 3.3%; Score 153.5; DB 2; Length 2366;	Query Match 3.3%; Score 153.5; DB 2; Length 2366;
Best Local Similarity 18.6%; Pred. No. 0.41; Mismatches 321; Indels 339; Gaps 51;	Best Local Similarity 18.6%; Pred. No. 0.41; Mismatches 321; Indels 339; Gaps 51;
Matches 186; Conservative 155; Gaps 51;	Matches 186; Conservative 155; Gaps 51;
27 NSALYFTNTSOPINDLQGSLAAYKFAQSQQLPQPHKEGSQPHLSLRLSRKSLLYRVPKAD	27 NSALYFTNTSOPINDLQGSLAAYKFAQSQQLPQPHKEGSQPHLSLRLSRKSLLYRVPKAD
RESULT 13	
A8312	hypothetical protein PA1874 [imported] - Pseudomonas aeruginosa (strain PA01)
C; Species: Pseudomonas aeruginosa	C; Species: Pseudomonas aeruginosa
C; Date: 15-Sep-2000 *sequence revision 15-Sep-2000	C; Date: 15-Sep-2000 *sequence revision 15-Sep-2000







X	(WISC ) WISCONSIN ALUMNI RES FOUND.			
X	Welch RA, Lathem WW;			
X	WPI: 2002-471441/50. N-PSDB; AAD38751.			
X	New pO157 plasmid-specified polypeptide found in Escherichia coli and other enterohaemorrhagic Escherichia coli, that binds to and cleaves C1-esterase inhibitor, useful for diagnosing and treating colitis -			
X	Claim 1: Page 52-56: 58pp; English.			
X	The present invention relates to novel pO157 plasmid-specified proteins found in Escherichia coli EDL933 and other enterohaemorrhagic E. coli, designated StcB, that bind to and cleave C1-esterase inhibitor. Sequences of the invention are useful for diagnosing, preventing or treating haemolytic uraemic syndrome or colitis in a subject infected with an enterohaemorrhagic pathogen expressing inhibitor protein. They are useful for testing a molecule for the ability to reduce proteolysis of C1 esterase inhibitor by inhibitor protein. The present sequence is E. coli O157:H7 StcE protein encoded by pO157 plasmid DNA.			
Q	Sequence 886 AA:			
Q	Query Match 100.0%; Score 4709; DB 23; Length 886; Best Local Similarity 100.0%; Pred. No. 0; Gaps 0; Matches 886; Conservative 0; Indels 0; Gaps 0;			
b	Y 1 MKLKLSTCLTILPLAIGVFSATAADNNSAIYNTSQINDLGSIAAETKFAQSOLPAH 60 1 MKLKLSTCLTILPLAIGVFSATAADNNSAIYNTSQINDLGSIAAETKFAQSOLPAH 60			
b	Y 61 PKEGDSQPHITSRSLLYRVPKAADDKTPVQEARDDNKKLIGPTLYPPSSLPDITYH 120 61 PKEGDSQPHITSRSLLYRVPKAADDKTPVQEARDDNKKLIGPTLYPPSSLPDITYH 120			
b	Y 121 LDGYPEGIDFTPHNCTKINTVAEVNKLSDAGSSSIHSHTLNNALEIHTANGRWRD 180 121 LDGVPSGIDFTPHNCTKINTVAEVNKLSDAGSSSIHSHTLNNALEIHTANGRWRD 180			
b	Y 181 IYLPGPDLSEKMRVFSAGYSSTVYFGDRKVTLFKYNGQWFRSGELENRR 240 181 IYLPGPDLSEKMRVFSAGYSSTVYFGDRKVTLFKYNGQWFRSGELENRR 240			
b	Y 241 ITYAOHIIWSAELPAHWIVPGNLNLVIRQGNLSGRNDIKIGAPGEELLHTIDGMITTPRD 300 241 ITYAOHIIWSAELPAHWIVPGNLNLVIRQGNLSGRNDIKIGAPGEELLHTIDGMITTPRD 300			
b	Y 301 RFDFAKDEAHREYFOTIPYSRMIYNNAPHLKEWMLPTGEELLTDMDPGNGGMWHSGTMR 360 301 RFDFAKDEAHREYFOTIPYSRMIYNNAPHLKEWMLPTGEELLTDMDPGNGGMWHSGTMR 360			
b	Y 361 QRIKGKLVSHGIDNANYGLNSTAGLGENSHPYVVAQLAAINSQRGNYANGLOVHGSSGGG 420 361 QRIKGKLVSHGIDNANYGLNSTAGLGENSHPYVVAQLAAINSQRGNYANGLOVHGSSGGG 420			
b	Y 421 IVTLDSTLGNEFHSHEVGHNTGLGHYDGFGRSVHRSAEINRNSTWDGDOKRFLPNFTPRD 480 421 IVTLDSTLGNEFHSHEVGHNTGLGHYDGFGRSVHRSAEINRNSTWDGDOKRFLPNFTPRD 480			
b	Y 481 QTNEKSCLNNOCQEPFDGHKEGFDAMAGSPSAANRFMTYTPNSSAIIOFFENKA VFD 540 481 QTNEKSCLNNOCQEPFDGHKEGFDAMAGSPSAANRFMTYTPNSSAIIOFFENKA VFD 540			
b	Y 541 SRSSTGFSKWNADTQEMEPYHTDRAQITASVNELESKMAELMAYAVKVHMGNG 600 541 SRSSTGFSKWNADTQEMEPYHTDRAQITASVNELESKMAELMAYAVKVHMGNG 600			
b	Y 601 WTRNIYIPTASADNRGSILTINHEACNSYLINGDEKYSQGKSFYSDGOFKERD 660 601 WTRNIYIPTASADNRGSILTINHEACNSYLINGDEKYSQGKSFYSDGOFKERD 660			
Q	Query Match Score 153.5; DB 17; Length 2366; Best Local Similarity 18.6%; Pred. No. 0; Gaps 339; Gaps Matches 186; Conservative 155; Mismatches 321; Indexes 339;			
Q	RESULT 2 AAR5011 ID AAR5011 standard; Protein; 2366 AA. XX AAR5011; XX DT 08-JUL-1996 (first entry) XX C. difficile toxin B. XX Toxin B; Cytotoxin; enterotoxin; fusion protein; antitoxin; XX Clostridium difficile VPI strain 10463 (ATCC 10463). XX PN WO9612802-A1. XX PD 02-MAY-1996. XX PF 23-OCT-1995; 95WO-US13737. XX PR 07-JUN-1995; 95US-0480604. PR 24-OCT-1994; 94US-032154. PR 16-MAR-1995; 95US-040496. PR 14-APR-1995; 95US-0422711. XX (OPHI-) OPHIIDIAN PHARM INC. XX PI Firca JIAMS, Kank JA, Padhye NV, Stafford DC, Thalley BS; PI Williams JA; XX WIPI: 1996-230603/23. DR N-PSDB; AAT29247. XX PT Fusion proteins comprising non-toxin protein and part of toxin - useful to form anti-toxins against Clostridium botulinum type A, and C. difficile type toxins, and to treat C. difficile intoxication, PT PT XX PS Claim 36; Page 313-323; 434PP; English. XX CC Clostridium difficile VPI strain 10463 toxin B (AAR5011), the product of the toxin B gene (AAT29247), is a cytotoxin associated CC with diarrhoeal disease. It can be obt'd. by expression in CC transformed E. coli hosts of portions of DNA that together cover the entire toxin B gene. Toxin B, and portions of it (see also AAR5012-13, CC AAR5371-72 and AAR5018), pref. expressed as fusions to Polynucleotide CC affinity tags or maltose binding protein, are used to raise anti- CC antibodies useful as antitoxins or diagnostics, and in vaccine prodn. XX SQ Sequence 2366 AA.			

Qy	27	NSAIYFNTISQPINDLOGSLAAEYKFAQSQILPAHPKEGDSOPHTSLRKSLLLVRPKAD	86	AC	AAW68388;
Db	1453	NSELOQNPKPYSVDFSEK-----ENGFIENGSTREG-----LFFYSELDPVLISKVMD	1500	XX	07-DEC-1998 (first entry)
Qy	87	DKTP-----VYEAARDDDNNKLIG-----TLLTLYPPSSLPTDTIYHLD-----	122	XX	Clostridium difficile toxin B.
Db	1501	DSKPSSFGYSSNNLKDVKVITKVNVLITGYLKDDIKISLTLQDEKTIKLNSVHLD	1560	XX	Antitoxin; vaccine; cytotoxin; toxin B; intoxication; immunogen; Pseudomembranous enterocolitis.
Qy	123	GPEGGIDFTPIHNGTKLINTAEVNKLSDAGSSSIHSHTNNALVEIHTANGRYRD-----	180	XX	
Db	1561	GVAE-----ILKFMRNKGNTN-TSDSLSMSFLESMNKSISIFNLFQSNIKFILDAN	1609	XX	Clostridium difficile.
Qy	181	-----IYLPGDPLEGKAYPRFVSSAGYSSSTVFYGDRTKVTL-----S	216	XX	W09B08540-A1.
Db	1610	FISGTSIGQFEEFICDENDNQPYFLKF-NFLETNTLYGNRQNNIVEPNYDLDGGD	1668	PD	05-MAR-1998.
Qy	217	VGTRLL---FKYNGQFERSGELEN---NRITYAQHWSAEL---PAH-----WIVPG	260	XX	EF 28-AUG-1997; 97WO-US15194.
Db	1669	ISSTVINFSSQKLYG-----IDSCYCNKVVISPNIYDTEINITYPEVTNTYPEVIVD	1721	XX	XX 28-AUG-1996; 96US-0704159.
Qy	261	LNLVIKGNNLCSRNLNDKI-----GAPGEVLLHTIDGMTTPRDRF-DEAKDKE-AHR	312	PA	(OPHI-) OPHIDIAN PHARM INC.
Db	1722	ANYNEKINNVN---INDSIRYIWSNDSDNDFLMSTSBNKVIKFVNVEKDTLANK	1779	XX	XX
Qy	313	EYF-----QTIPYPSRMVNNYAPHLKEVML-----	338	PI	Thalley BS, Williams JA;
Db	1780	LSEFNESDKQDVSEIIL-SFTPSYXEDGLSYDLGLSYLNEKEYINNFCHMVSGLIYI	1838	XX	XX
Qy	339	339 -----PIGBLLTD-----MDPGNGWHSRGTMRIGRELVSGIDNANYGL	379	XX	Host cell containing recombinant expression vector encoding Clostridium botulinum type B or E toxin - useful to treat humans and other animals at risk of intoxication with clostridial toxin
Db	1839	NDLSLYYFKPVNNNLITGFYTGVDDKKYFNPINGGAAS-----IGETI---IDDKRNYF	1888	XX	WPI: 1998-230234/20.
Qy	380	NSAFLGGENSHP-VVAOLAAHNSRGNT-----ANGIQHGSGGGCIVTDLSTLGNE-----	431	XX	Example 18: Page 241-249; 428PP; English.
Db	1889	RQG-----VLTQGTVFSTEDGPKYFAPAN-----TLDENLEGAI 1924	XX	XX	
Qy	432	FSHEVGHNYGLGHYVDGFKGSYVHRSAAENNNSNWGW-----DGDKKRFLPNFYSQOTNEKSC	488	CC	This is the amino acid sequence of Clostridium difficile toxin B, deduced from the coding region (see AAW30561) of the toxin B gene.
Db	1935	FTKLIIDENITYFDDAYTRGAVE-----WKEDGEMHYFSP-----	1961	CC	Fragments of the toxin B gene have been cloned into various prokaryotic expression systems, and assessed for the ability to express recombinant toxin B protein in E. coli. It would be advantageous to use simple and inexpensive prokaryotic expression
Qy	489	NNOCQEPDPDGHHKEFGFDAMAGGSPSAANRF-----TMYTPNNSAIIORFF-----ENKAVFDSR	542	CC	systems to produce and purify high levels of recombinant toxin B for immunisation purposes. The invention specifically relates to recombinant proteins derived from Clostridium botulinum toxins (see AAW68389-400) and their use as immunogens for the production of vaccines and antitoxins.
Db	1962	-----TGKAFGLNQIGDKYTFNSDGTMQKGTVSINDNKHYFDD	2002	CC	XX
Qy	543	S-STIGFSK-----WNADTOEHEPEYTIDRAEQITASVNESEMAIAMEAYAVKV	594	Sequence	2366 AA;
Db	2003	GPKVGYCTEGDKHFAENGEMQIGVNTEDGFKYFAHNEDLGNEEGEISYSGIL-----	2060	Query Match	3.3%; Score 153.5; DB 19; Length 2366;
Qy	595	HAWNGNMTTRNIYIPTASADNRGSILTINHEAGYNSYLFDKVVSGQYKKSFSVSDQF	654	Best Local Similarity	18.6%; Pred. No. 0.033;
Db	2061	-NFNNKYY-----FDDSFATAVGWDLEDGSKYF-----DEDTDAAYIGSLINDQY	2109	Matches	186; Conservative 155; Mismatches 321; Indels 339; Gaps 51;
Qy	655	WKERDWDVTREARKPFQFG-----VPTVTLGYDPEGTLOSSIYIAMYGAYGFTYSSDDSONL	713	Qy	27 NSAIYFNTISQPINDLOGSLAAEYKFAQSQILPAHPKEGDSOPHTSLRKSLLLVRPKAD 86
Db	2110	YFNDDGI-----MOVGFVTFINDKTFYFSDGIIIES-----GYQNIDDDNYFYI	2151	Db	1453 NSELQKNPKYPSVDFSEK-----ENGFPNGSTNEG-----LFTVSELPDVLSKVMYD 1500
Qy	714	DNDQOLCY-----DTKEGQRLRANHRAANTVANKFHINVPTESOPTATLVCNNKILDTK	770	Qy	87 DKTP-----VQYEARDDNNKKLG-----TLTLYPPSSLPLDTIYHLD-- 122
Db	2152	DDNGIVQIGFDTSDGKYYFAPA-----NTVRDNII-----GOAVEYS	2189	Db	1501 DSKPSFQYSSNNLKDVKVITDNVNTLGYLKDDIKISLSLTDQETKILNSVHLD 1560
Qy	771	SUTPAP-----GLTYTYNGQALPAKENEGCIVSVNSGKRCCLPVGQRS-----GYSLSDPWIV	823	Qy	123 GPVEGGIDFTPHNGTKLINTVAYENKLSDAGSSSIHSHTLNNALVEIHPANGRWRD-- 180
Db	2190	GLYRVEDYIYFGETYIETGWIYDME-----SDKYYINPETKACRGINLDDI-----	2241	Db	1561 GVAE-----TSDSIMSPLESMNIKSVTFNLOSNIKFTLDAN 1609
Qy	824	QOEYVYDSGAKAVYLLSDMDLSY-----NRIGEF-VGNNVPAD	861	Qy	181 -----IYLPQGPDEGRMRVRFSSAGISSTVFYGDRTKL-----S 216
Db	2242	-KYYFDERGIMTGLISFENNNYYFENNGMFGYIINIED	2280	Db	1610 FLISGTTSIGOFEFICDENDNQPYFKF-NYLENTNTLYGNRNDMIVFEDLDSSGD 1668
Qy	261	LNUVIRGQNLCSRNLNDKI-----GAPGELLAHATDIGHMTTPRDRF-DPAKDRE-AHR	312	Qy	217 VGNTLL-----FKYVNGQFERSGELEN--NRITYAQHWSAEL--PAH-----WIVPG 260
RESULT 3 AAW68388 ID AAW68388 standard: Protein: 2366 AA.				DB 1669 ISSTVINSQKLYG-----IDSCYCNKVVISPNIYDTEINITYPEVIVD 1721	

1722	ANYINEKINNN	-INDSIRYTWNSNDNDFILMSTEENKYSQVKLKFVNFKDK2LANK	1778
313	EYF-----QTIPVSRMIYNNYAPHLKEYML	-----	338
1780	LSFNFSRQDVPVSEIII-SFTPSYEDGLIGDGLVSLYNEKFYINNFGRHNVSSLIY	1838	
339	-----PTGELLTD-----MDPGNGWHGSTMQRIGKELVSGIDNANANGL	379	
1839	NDSLYYFKPVPNNLITGFVTYGDKXYFNPNTNGAAS-----IGETI-----IDDDNNYYF	1888	
380	NSTAGLGENSEHPVVAOLAHHNSRGNY-----ANGIYHGCGGGIVTLDLSTLGE-----	431	
1889	NQSG-----VLTQTCVGFSTEDGPKYFAAN-----TLDENLEGAIID	1924	
432	FSEHEGHNYGLGHYVGDGFKGSVRSAENNNSTMGW-----DGDKKRFLPNFYPQSQTNEKSCL	488	
1925	FTGKLIIIDENIYFFDDNTRGAVE-----WKELDGERHYFSP-----	1961	
489	NNQCQEPDGHKGFDAMAGGSFSAAFRF-----TMYTNTSSAIIQRF-----ENKAFTDSR	542	
1962	-----TGKAFKGQLNQIGDKYKVVNSDGYMQKGTVSINDNKHFFDDS	2002	
543	S---STGFSSK-----WNADTQEMPEYHTIDRAEQITASVNELESKMAELMAEYAVKV	594	
2003	GTVKYGITIEDGKHFYFAENGEMQIGVNTEDCFKYFAHHNEDLGNEEELSYSGSIL-----	2060	
595	HWWNGINTRNIYIPTASADNRGLILTINHEAGUNSLFINGDEKVKSYOGKSFVSDQF	654	
2061	-----NFNNKIIY----FDDDSFTAVGWKDLEDGSKYIF-----DEDTAAYGISLNDGQY	2109	
655	WKEKDDWDTREARKPEQFG-VPTVTLGYYDPGTLSSXYPAMKGYGFTYSDDSQNL	713	
2110	:YENDGII-----MOGVFTVLTNDKVFSDSGIIES-----	2151	
714	DNDQCQLQV-----DTKEQGLRFLRNHRRANTVNMKFHINVPTESQPTQATLVCNKKLDTK	770	
2152	DDONGVQGFQDFSDGYYKFAPA-----NTVNDNIX-----	2189	
771	SLTPAPE-----GLTYTVNGQALPAKENEGCIVSVNSSERKYLPGORS-----GYSLPDWIV	823	
2190	GLYRGVEDVYYFGETTYIETGWIYDMENE-----SDKYYFNPETKACKGINLDDI	2241	
824	GOEYPLWDSCKAKYVLLSDMDNLSY-----NRIGEF-VGNTNPAD	861	
2242	-KYYFDERGIMRTGGLISPFENNYYFNENGEMOFQYINIED	2280	
RESULT 4			
D	ABG30537 standard; Protein; 1249 AA.		
X	ABG30537;		
C	07-OCT-2002 (first entry)		
X	Alpha-isomaltosylglucosaccharide synthase #1 mature protein.		
W	Alpha-isomaltosylglucosaccharide synthase; sugar production;		
W	cyclic tetrasaccharide; cyclo(-6)-alpha-D-glucopyranosyl(1-3);		
W	alpha-isomaltosyltransferase; viscosity modifier; sweetener;		
W	heat stabiliser; hydrolysis stabiliser; colour stabiliser; foodstuff;		
W	cosmetic; drug composition; enzyme.		
S	Unidentified.		
S	WO20255708-A1.		
X	18-JUL-2002.		
D	09-JAN-2002; 2002WO-JP00052.		
F	12-JAN-2001; 2001JP-0005441.		

A S Unidentified.  
X X  
N WO200255708-A1.  
X X  
D 18-JUL-2002.  
X X  
F 09-JAN-2002; 2002RG-JP00052.  
X X  
R 12-JAN-2001; 2001JP-0005441.  
X X  
A (HAYB ) HAYASHIBARA SEIBUTSU KAGAKU.  
X I Kubota M, Maruta K, Yamamoto T, Fukuda S;  
X I  
X X  
R WPI; 2002-520129/55.  
R N-PSDB; ABR88157.  
X X  
X X New alpha-D-isomaltosylglucosaccharide synthase of bacterial origin for

The invention describes novel microbial polypeptides having alpha-1-sialylglucosaccharide synthase activity. The proteins are useful for producing a sugar (I) having at least three glucose units connected by alpha-1,4 bonds except for alpha-1,6 at the non reducing end. The invention also describes a method for the production of the cyclic tetrasaccharide cyclo(-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-glucopyranosyl(1-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-glucopyranosyl(1-2) (II) by treatment of (I) with alpha-L-sialic acid transferase. (II) and similar sugars in crystalline or syrup form are used as a viscosity modifier, sweetener, heat stabiliser, hydrolysis stabilisers and colour stabiliser for foodstuffs, cosmetics and drug compositions.

CC	XX	Sequence	1284 AA;
CC	CC	Query Match	3.28; Score 149; DB 23; Length 1284;
Best Local Similarity	18.8%	Pred. No.	0.028; Mismatches
Matches	170;	Conservative	105; Mismatches
Matches	170;	Indels	308; Gaps
Qy	105	TITLYPPSSLPDITYLHLDGYEG--GIDFPHNGTAKIINTVAEYNKLSDASGSIIHSHH-	161
Db	54	TITDVQNSAEPDDLLVQAVONGILRVDYRNPNTSFAKTPMLDPKTKTWSAVGATNTA	113
Qy	162	-LTNNALVEI-----HTANGRWYRDYLPOGPDELSKMRVRFSSAGYSSTVF	207
Db	114	NPMTTISNNKRIEITKRPVTRVKAQDGTTL--FWPSGGGVSFSDGRFLRHFATGDN--M	168
Qy	208	YGDRTKTVLTSVGNLTLKYVRSWTFRSELENNRITAQH!WSAELAHWVPGLNIVIKQ	267
Db	169	YGIRSF-----NAFDSSGDLRNNSNNRAAH-----	193
Qy	268	GNLSGRLNDIKIGAPGELLALHTIDICMLTTPRDRFEAKDKEAHRPFQ--TIPVSRMIV	325
Db	194	--AGEGDD---SGGPLIWSFTAGYGLLVSDGGYTDTSTGTQMMFYGGTPPBRRYA	245
Qy	326	NNYAPLHL----KEYMLPTGEITMDPGNGWHSGTMR-----	361
Db	247	KONVEYIMLGSTPKPELMIDYGE-ITGKPPMLPKWSMGFMPEWDNTQEFNNVDTYRAK	305
Qy	362	RIGKELYSHGIDNANTGLNSTAGLGENSHPYVVAQLAHNRSNGVQHG-----	414
Db	306	NIPIDAFAFDIWKKGETNYGEFANTNTNEPSASTSLKSTD-AKGKIRMIGITPKRV	364
Qy	415	-GSGGGGIVT--LDSTLGNFESHEVGHENYGLGHYVDFKGSYHRSAE--NNNSTNGWDG	468
Db	365	TDASANVTTQTDATANGIF--YPGIN--EYQDFIPTVRSIDPYNANERAFWNH	418
Qy	469	DK--KRIFPNFVPSQTNEKS-----CLNNOCQEFPDGHHFGFDAMAGGSP--	511
Db	419	STDALNRKGIVGMNDTQDKVSSGGALIYWFQNETTGIMQSOTMEEGGR--AYTSGQRYW	474
Qy	512	--TSAANR--TAYTP-----NSATQRFQENFDSSASSTGGS	551
Db	475	QTARTFFPQGQRYATLWNSDIDQIYNGKERIWNAGCMQ--EQRAVMFLSYANGQVKWG	531
Qy	552	ADT-----OEMEPY-----ERTIDRA	567
Db	532	MDTGGFNCQDGTINNPFDLYARWQOISALTPVFRVHGNHQCQRPWYFGSTAEASKEA	591
Qy	568	EQI-----TASYNVELSESKMAELMARY--AVVK--VHMNN-GNW-----	601
Db	592	IQLRSLIPIKYAYERAYENGILVRPLMQAYPTDAAVKNTDARKMFQDNLLAAPVYDK	651
Qy	602	--TRNIYIPTASADN--RGSLIT-----INHEAGYNSYLFINDEKEVVSOGJJKSF	648
Db	652	QQTSKDIYLPSCSWIDYARGNMTGOTIRSYNPDLTDAPLFI-----KKGA	700
Qy	649	VSDGQFWKNERDUDTREARKPFGPVPTLVG--YKDPEGTLSSLYPAMYGAYGFTYS	706
Db	701	LIPTQ--KVQDYVGASVTSVYDVFDDTQSSTYDDG-----ASYNYESGTYF	750
Qy	707	DOSQNLSDNDCOLQDVTKEGQFLFRILHRNNTYANKFHINVPTSQOPTATLYCQNKKI	766
Db	751	KONMATQDNG-----SGSLSFTGAKSGSYTPALQSYI--VFLHGSAGTSVNNNS-	798
Qy	767	LDTKSLTPAPEGTYTNQGOALPAKENEGCTVSVNSSKCRYCLPVGGRSQSYLSPDWVYQOE	826
Db	799	--AAMTSIASLEAKAAGEG-----WATGRD	823
Qy	827	VYVD 830	
Db	824	IVYD 827	



Production of isomaltose used for manufacturing e.g. beverages, health foods and cosmetics by using specific non-reducing end alpha-1,4-glucosyl linkage-containing saccharide and alpha-isomaltosyl glucosaccharide synthase - Disclosure; Page 103-111; 121pp; Japanese.

The invention relates to a novel method for the production of isomaltose, comprising treating a saccharide having alpha-1,4-glucosyl linkage at the non-reducing end and a glucose polymerization degree of up to 2 with alpha-isomaltosyl glucosaccharide synthase, treating the product(s) with isomaltose-releasing enzyme and collecting the obtained isomaltose. Isomaltose is useful for manufacturing beverages, health foods, feeding materials, cosmetics, drugs and cooking sauces. The sequence represents a polypeptide used in the invention. No further information about the sequence can be obtained from the specification.

Q	A	Sequence	1284
		Query Match	3 - 2%
		Best Local Matches	18 - 8%
		Local Similarity	105 / 105
		Conservative Matches	321: Mismatches
		Pred. No.	0.028;
		Gaps	428;
b	b	105 TLTLKPPSSLDPDTIHLGDVPEG --GIDFTHPGTKIINTVAYENKLSDASGSFHSH - 161	
b	b	54 TLTVNDGAEPDDLLIVQAVQNGLKVADYRNTSITPSAKTMLDPKMTKWSAVGATINTTA 113	
b	b	162 -----LTNNALAEI-----HTANGRWYRDIVYLPOGPDLLEGKAVRFVSSAGSTVTF 207	
b	b	114 NPMTTISNMKIEITTNPVATVKADGTL - FWPSGGGVTSDFGVRFLATGDN - -M 168	
b	b	208 YGDRKVTLSVGNTLFLKYVNGQWFRSGELENNRITYAQHITSAELPAHWIYPGLNLVYIKQ 267	
b	b	169 YGIRSF-----NAFDSCGDLRNNSNHAH----- 193	
b	b	268 GNLSGRGLNDIKIGAPGELLNTIDIGMLTTPDRDFEAKDEKAHREYFQ - TIPYSRMIV 325	
b	b	194 ---AGRGD---SGGPLINSTAGYGLYLDSDGGYPTDSTGTQMFYGGTPPERRYA 246	
b	b	326 NNAYPLHL-----KEVMLPTGELLYTDMDPGNGGMHSGTMR-----Q 361	
b	b	247 KQNVYEYIMLGTPKEIMDVG - ITKGKPPMPKWSLGFMMNFWDTGTOETFNNVVDYRAK 305	
b	b	362 RIGKELYSHGIDNANGLNINSTAGLGENSHPYVVAOLAHHNSRGNYANGIQVHG - - - - 414	
b	b	306 NIPIDAYAFDYDWKYGTEINYGEFAWNNTTNPFSASITSLKSTMD - ARGIKAGITKPRIV 364	
b	b	415 -GSGGGIVT - LDSTLGNFESHEVGHNYLGHYDGFKGSYHRSAE --NNNSTGWWDG 468	
b	b	365 TKDASANVTQGTDATNGGYF - YPGHN - -EYQDXFIPYTVSYTQDYNANERAWFWNH 418	
b	b	469 DK -- KRFIPNFPYSOTNEKS -----CLNNOCQEPFDGKFGFDAAGMGS --- 511	
b	b	419 STDALANGTVGWNDEDKVSSEGALYWFGNFTGHSQMTVEGGR --- AYTSGARVW 474	
b	b	512 ---- FSAANRF - TMYTP -----NSALIQORFFENKAVFDSRSSTSFGSKW 551	
b	b	475 QTARTYFPGAQRYATLWSGDIQYOKNGKERINWAAGMO --EQRAYMLSSTYNNQGKWF 531	
b	b	552 ADT - -----QEMEPY -----ERTIDRA 567	
b	b	532 MDTGFGNOQDGTTNNPNPDLYARMQFOSALTLPVFRVGNHOOQPYWFGSTAEASEKA 591	
b	b	568 EOI - -----TASYNELSESKMAELMARY - -AVK - VHMNP - GNW ----- 601	
b	b	592 IOLRYSLIPYMTAYERSAYENGNGVRPLMQRQYPTDAVKNTTDAWMGDNLLAAPYVDK 651	
b	b	602 --TRNVIYIPTASADN - RGSILT -----INHEAGNSYLFINDEKVSGQYKKSF 648	
b	b	652 QQTSDKDIYLPGSSWIDYARGNAITGGQTIRYWSNPDTLTDMPLFI -----RKGA 700	
b	b	649 VSDGOFMKERDVDTREARKPEOFGYPUYTLYG - -YDPEGTLSXYTAMYGAGFTYS 706	

701 RIPTQ -KVQDYGQASVTSVDYDFDTIQSSFTYYDDG-----ASYNTESGYTF 750  
 Db : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 707 DDSQNISDNDCQLOQYDTKEGOLFRLAHBRANNTWINKFHINVPTEQPTQATLVCNKKI 765  
 Qy : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 751 KONRATHQDG-----SSESILIGAKSGSYPPALQSYI---VKLHSAGTSYNNNS 798  
 Db : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 767 LDTKSUTPAPEGLTYVNGOALPKANEKGCTIVSYNSGKRYCLPVQRSGYSLPDMYQE 826  
 Qy : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 799 -----AAMTSVASLEALKAAAGEG-----WATGRD 823  
 Qy : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 Db 827 VVVD 830  
 Qy 824 IYGD 827

**RESULT 8**  
**ABB47329**  
**ID ABB47329 standard; Protein; 1530 AA.**  
 XX  
 AC ABB47329;  
 XX  
 DT 05-FEB-2002 (first entry)  
 XX  
**Listeria monocytogenes protein #33**  
 DE  
 XX  
 KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 KW vitamin B12; bacterial infection; disease.  
 XX  
 OS Listeria monocytogenes.  
 XX  
 PN WO200177335-A2.  
 XX  
 PD 18-OCT-2001.  
 XX  
 PR 11-APR-2001; 2001WO-FR01118.  
 XX  
 PR 11-APR-2000; 2000FR-0004629.  
 XX  
 PA (INSP ) INST PASTEUR.  
 XX  
 PI Buchrieser C, Frangoul L, Couve E, Rusniok C, Fslhi H, Dehoux P,  
 PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P,  
 PI Daniels J, Goebel W, Kreft J, Ruhn M, Ng E, Vazquez-Boland JA,  
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martin A, Amend A,  
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L,  
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N,  
 PI Madueño E, De Pablos B, Weiland J, Kaerst U, Entian K, Hauf J,  
 PI Rose M, Voss H,  
 XX DR WPI; 2002-010914/01.  
 XX PT Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment  
 PT and prevention of *Listeria* and related bacterial infections, and  
 PT related polypeptides -  
 XX PS Claim 6; SEQ ID No 34; 192PP; French.  
 XX CC The present invention relates to the genome sequence of *Listeria*  
 CC monocytogenes EGDe (see ABB03041). The genome sequence and fragments of  
 CC it are useful for selecting probes and primers for detecting genes in *L.*  
 CC monocytogenes and related organisms, and for studying genetic  
 CC polymorphisms and other genomes. The present sequence is a protein  
 CC encoded by the genome sequence of the present invention. Proteins  
 CC expressed from the genome sequence are useful for raising specific  
 CC antibodies, identification of *L.* monocytogenes and related organisms, and  
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
 CC B12. The genome sequence and proteins encoded by it are also useful for  
 CC selecting compounds that regulate gene expression and cell replication  
 CC and modulate *L.* monocytogenes related diseases. In addition, the genome  
 CC sequence and proteins encoded by it are useful in pharmaceutical and  
 CC monocytogenes compositions for the treatment or prevention of infections by *L.*  
 CC monocytogenes and related organisms.

Note: The sequence data for this patent did not form part of the printed CC specification, but was obtained in electronic format directly from WIPO CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

Sequence	1530 AA:
Query Match	3.1%; Score 147.5; DB 23; Length 1530;
Best Local Similarity	19.9%; Pred. No. 0.049;
Matches	205; Conservative 136; Mismatches 348; Indels 341; Gaps 59;
Qy	17 GFVFSATAADNNTATFNTSQPI---NDLQGSLAAEVKAQSSQI-----LPHP----- 61
Db	597 GVFNYKTGELEWKLFIDNSKSLISKPTIEDLSLUSGOTIQDSIEHKIDLSATPQVCELI 656
Qy	62 -----KEGDSOPHLTSRKSLILYRYPKADDKT-PV----- 91
Db	657 PENTYDVTFTKPKENGNEQMLTFKKP--LIHP/EVTYTKPKGITKPLPKNAVISDGE 714
Qy	92 -----QEARDNDN MKLGLTLYPPSPUDTYYLHDQYPEGFIDFPHNGTKKINTVA 145
Db	715 VLADYEAEVDNANKYVAKRSGEQYGDNDWETY---ANQS---STVSNAF---VTPDLG 766
Qy	146 EVNKLDASGSISIHSHTLNNAALVEHTANGRWTDIYLQPQP-----DLEGK 192
Db	767 TGQLR-DTSS1KYKSQT----SVTKMQLQESNMPISEYDILKTGVDEESNLXF 817
Qy	193 MVRVSSAGYSTVY-----GDRKVTLSVNTLLEFKYVNSQWFRGELNNRITYAQ 245
Db	818 QVKXKNEINOSVYKYQTAITLISDTETAQNSVTF---TDNITKGETEKTKE---- 669
Qy	246 HIWSAELPAHWWIVPGNLVYKOGNLSGRNLKDIGAPGELLHTIDGMLTTPRD--RFD 303
Db	870 -----NIEVKITGTGDT----TGETGKILANKVDAKDPSPLEGATED 909
Qy	304 -FAKDEKAHREYFQ--TIPVSRMIVNNYAPLHKEYMLPTEGELLTMDNGGGWHSGTM 359
Db	910 LYANDEKVDTQTTDKNGVIEFDLILYGYDT---LKEVASPEGTLPTASTEN-----I 959
Qy	360 RQRICKELVSHGIDNANYGLNSTGLGENSHPTVQAQLAHNNSRGNYANGIQVHGG--- 415
Db	960 QYKLEQDEKTVYQVANEKMPKET--GE-----VHLVKTDKATGATLAGAEFSL 1005
Qy	416 ---SCG---GGIVLDSLGEFNEFHSHEVGINYGLHV-----DGKF---- 450
Db	1006 YDKSAELQNLTT----DENGELTINLNDLGSYSLKETKAPEGYKLSEKZWESYES 1059
Qy	451 ---GSYHRSAENNNTSWGMDGDKKRFPNFYPSOTNEK-----SCLNN----- 490
Db	1060 GQVDALEIQAN-----ERDGEAVLTKVDSSETNAKUGAFNLNDSGEV10TNLVS 1112
Qy	491 ---QCQEPPDGHKFGDAMAGGSPPSAANR--ETMYTPNNSALLQRFENKAVP 539
Db	1113 DENGELTRVQNLQEPGD---TAQQTETAPNTNDLATNTWPPTIVAGQTATMVAENK--- 1166
Qy	540 DSRSSTGFSRMADQEMEPYERTIDRABOITASVNESESMAELMAYAVVVRHMRG 599
Db	1167 ---TG--KPDVDTGEVILVKQDSATGETLEGAVDL----MTADGATVASNLL--- 1209
Qy	600 NWTNRNVIYPTASADNRGSLTINHEAGYNSYLFINGDEKYSOGYKSFVSDQFWKERD 659
Db	1210 -----ITDANGEITVTVNLAPGKYSF----KETKAPEGYELA-----TD 1243
Qy	660 VWD-TREARKEPEQFGYPV-TLVGYDPEGTLSKYYIYPMYGAFTYSDDSON----- 711
Db	1244 WFEFTIAPNOPEKTTITAENTKLAPIPDAGSVK-----IIQDSENGVRLAG 1290
Qy	712 ----LSDNDQCOLQVDTK---EGQFLF-RLA---NFRANNNTVMKFEHNPTESSOPTQATL 760
Db	1291 AEFSLIAENGETLQTMKTDGEALEVNLLAPGNRYIQET---KAPDXQLESTPWFEL 1347
Qy	761 VCNNKNDLTKSLTAPPEGLTYTVNGOALPAKENECIVSVNSGKRYCLPVGQR-SG--YS 817
Db	1348 VAN----DTISCVTVAE-----NAKLEPDVAYTCAVLKTDSE---TGTRLSGAVFS 1393

Qy 818 LPDWIVGQEVYVDSGARA KVLLSDMDNLSYNRIGE-FVGNVNPAD--MKRVKAVNGQVYLD 874  
 Db 1394 LLD-----ESG---KVQA---NLTDENGELFIDGDLTPGNSLKETKAPGYEL- 1437

Qy 875 FSKPRSMRVV 884  
 Db 1438 AEQPWNFQIV 1447

RESULT 9  
 ABC30539 standard; Protein; 929 AA.  
 XX  
 AC ABG30539:  
 XX  
 DT 07-OCT-2002 (first entry)  
 XX  
 DE Alpha-isomaltosylglucosaccharide synthase #3 mature protein.  
 DE Alpha-isomaltosylglucosaccharide synthase; sugar production;  
 KW cyclic tetrasaccharide: cyclo(6)-alpha-D-glucopyranosyl(1-3);  
 KW alpha-isomaltosyltransferase; viscosity modifier; sweetener;  
 KW heat stabiliser; hydrolys stabiliser; colour stabiliser; foodstuff;  
 KW cosmetic; drug composition; enzyme.  
 OS unidentified.

XX  
 PN WO200255708-A1.  
 XX  
 PD 18-JUL-2002.  
 XX  
 PF 09-JAN-2002; 2002WO-JP00052.  
 XX  
 PR 12-JAN-2001; 2001JP-0005441.  
 XX  
 PA (HAYB ) HAYASHIBARA SEIBUTSU KAGAKU.  
 XX  
 PI Kubota M, Maruta K, Yamamoto T, Fukuda S:  
 DR WPI: 2002-520129/55.  
 DR N-PSDB; ABR8156.

XX  
 PR New alpha-isomaltosylglucosaccharide synthase of bacterial origin for  
 the production of cyclic tetrasaccharide gum -  
 PT Claim 1; Page 109-112; 144pp; Japanese.  
 PR XX  
 CC The invention describes novel microbial polypeptides having  
 alpha-isomaltosylglucosaccharide synthase activity. The Proteins are  
 useful for producing a sugar (I) having at least three glucose units  
 connected by alpha-1,4 bonds except for alpha-1,6 at the non-reducing  
 end. The invention also describes a method for the production of the  
 cyclic tetrasaccharide cyclo(-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-  
 glucopyranosyl(1-6)-alpha-D-glucopyranosyl(1-3) (II) with alpha-isomaltosyltransferase,  
 (II) by treatment of (I) with alpha-isomaltosyltransferase,  
 (II) and similar sugars in crystalline or syrup form are used as  
 a viscosity modifier, sweetener, heat stabiliser, hydrolysis stabiliser  
 and colour stabiliser for foodstuffs, cosmetics, and drug compositions.  
 This is the amino acid sequence of an alpha-isomaltosylglucosaccharide  
 synthase mature protein.

XX  
 SQ Sequence 929 AA:

Query Match Score 145; DB 23; Length 929;  
 Best Local Similarity 19.8%; Pred. No. 0.035;  
 Matches 165; Conservative 107; Mismatches 288; Indels 272; Gaps 49;

Qy 176 RWVRDIXYIPOGQ---PDLERGAKYRFVSSAGYSSSTVYDFDRKVTLGVNTLLEFKVNGQWFR 232

Db 210 RW---YLMGGDLDPLDRHDYMELTGTPPPVPPKKAFG---LWVSE---FGYDN--W- 253

Qy 233 SGELLENRRNITYAQHWSAELPAHIVPCLNLYIKQ-CNLSGRNLNDIKIGAPCELLLHTI- 290

Db	254	-SEVDNT----IAGLRSADDFP----VDAAMLQDYQWFGGTTAIDSDTRMG-----TLD 296
Qy	291	-DIGMELTPDREF-DAKDKDEAHRETFOTIPYSRMIVNNYAPLHLKEYALPTGEILTD-- 346
Db	297	WDTSRFPDPAGKIAADLADFGV----TPIPESSYGNLPEHAR--HAADGYLYRSGC 348
Qy	347	----MDP----GNGG-----WHSCTMRQIGKELYSHGIDNAN--YGLN 380
Db	349	ATCPPVLTGPNPWKGGMIDWTOPEAQAVWHDEQRQLHVDEGVIGHWLDLGEPERMDPN 408
Qy	381	S-TAGL----GENSH-----RTVVAQLAHNRSRGYANGI 410
Db	409	DWTAGVIPGKHADYHNAVNLWAQSIADGYADNGVQRPFMLTRA-----AAGI 460
Qy	411	OYHG-GSGGGGIVYLDSLTLGNFESHEVGHNW-GLUHY--VDGPKGSVHRSAAENNINSTWG 465
Db	461	QRHGAGAMNSADIGSTMKLGSSQNAQMIMSMGDDYGSDDIGGRREM-ADGDVNELYTQ 519
Qy	466	WDGDKKREKIPNFYPPSQTNEKSCUINNOCQEPFDGHKFGFDAAMGGPSAANRFTMTPNS 525
Db	520	WFDSAEDTPLRPHTDNLNCLET-----SPDSICD-----VASNR 556
Qy	526	SATQRFENKAYEDSRSSGTGFSKWNADTOEMEPYEHITI-DRAEQI----- 570
Db	557	ENLVRY-----ELAPPYYSLAHRAHGFEGEPLAPPLYQQN 593
Qy	571	TASVNELESKMAELMAEAYAVVKYHMNGNWTRNXYIPTASADNRGSLTINHEAGYNSY 630
Db	594	DDIVREVMGHQKM---LGDRLLIAVAGEGEERDYLP-----GEW 632
Qy	631	LFLINGDEKVVSQGKYKSFYSDGQTKWERDVDTREARK----PEQFGYPVTIILYYD 684
Db	633	IDIHTNERIQTSGT---QWIDNPVLW--RGQVFTLPAYARAIIFFKAFAVDASTKDITGKR 687
Qy	685	PEGTLSSTYIPAMYG---AYGFT-YSDDSQNLSNDQCOLQYDTKESQRLFRLANHRANT 740
Db	688	EDAVRNELIATYADDVASDTILEDDDTAYADGAVRTT---QISQSLTNQGVATVT 743
Qy	741	V--MNKFH1INVPTESOPTQATLYCNKILDTKSUTPAPEGLTYTNGOALPAKEN---- 793
Db	744	VGAASGTYSGAPS-TRPTVVELDTGTQASTVSUGSVP--LTEHANKAAFDAAQSWYNA 800
Qy	794	-EGCIV----SYNSGKRYCLPGVQSGYSL----PDIVGQEVYVDSGAKAVVLLSD 841
Db	801	GGGLIVVAKAAASSSYNTAKTFSFTLGEESTWATFSENATTFGQSYYV----- 848
Qy	842	WDNFSLYNRIGEF---VGNNVPADMKVKYK----ARNQYLDFSKPSRMSMRYVYK 886
Db	849	-----VGNVPOLGNNSPADAVKLEPSAYPTWTGVVRNL--PPSSTVENK 890
	RESULT 10	
	ID	ABG30565 standard; Protein: 965 AA.
	XX	
	AC	ABG30565;
	XX	
	DT	07-OCT-2002 (first entry)
	XX	
	DE	Alpha-isomaltosylglucosaccharide synthase #3.
	XX	
	KW	Alpha-isomaltosylglucosaccharide synthase; sugar production;
	KW	cyclic tetrasaccharide; cyclo(-6)-alpha-D-glucopyranosyl(1-);
	KW	alpha-isomaltosyltransferase; viscosity modifier; sweetener;
	KW	hydrolysis stabiliser; colour stabiliser; foodstuff;
	KW	cosmetic; drug composition.
	XX	Unidentified.
	OS	
	XX	
	PN	WO200255708-A1.
	XX	



Qy	772	LTPAPELTYVINGQALP-AKEENGCG-----IVSYNSGKRYCLPVGQRSGYSLPDWYG	824	Db	12	LSVLAASALIGSTVPSLAPPQAHVSSAIGNLSSAYTGTDLTLTIDNGAEPNDDILVHQ	71
Db	1775	-GMADYNGQHVVQYKOGDAYAQRLITLNN-----	1806	Oy	123	GYPFG--GIDFTHPHNGTKRINTVYAEVNKLDSLGSSTHS----HLJNNALEIHTANG	175
Qy	825	QEYVDSGAKAVLSD-----W-DWNLSYARLIG-BEVGVNNYPAADMKKVKA	867	Db	72	AVONGILKVDYRPNGVAASADTPMLDPNKTWPPIGAVINTASNPMTI-TTPAM-KIEIAKN	130
Db	1807	QOVWTDSSLRALSSTIMOAMNDMVYNSNSORTDGLWLNAPIYMSGAKWAGDTSA-----	1859	Oy	176	RWVRDYLFLQG-----PDLEGK-----WFRVSSAGYSSTVFYCDRKYTTSVGNLTLEFKYV	225
Qy	868	WNGOVLDFSKPRSMYR 883	1860	Db	131	PVRLYTKDGTALLWEPPGGYVESDGTREFLHGTDN--MGIURSFNAEDSGDULRNS	187
Db	1860	-NGRYVHISKAYSNEV 1874	1888	Oy	227	NGQUNFRSGELENNTYAOHIWSAELPPAHWIVPGLNLVYIKOGNLGGRNDIKIGAPGELL	286
				Db	188	STQARAGD-----QGN-SG-----GPLI 205	
RESULT 12				Oy	287	LHTIDIGMLTTPDRDFARDKEAHEYFQ--TIPVSRMIVNNYAPLHL-----KEYMLP	339
ABG30564	ID	ABG30564 standard; protein: 1286 AA.	XX	Db	206	WSTAGYGVLYVDSOGYPTDEATGKLEFYGGTPPEGGRTKQDVETIIMLGTPKEIMSG	265
AC	ABG30564;		XX	Oy	340	TGELLTDMDPGNGSGWHSGSMR---ORICKELVSH-----GIDNANTYGLN	380
DT	07-OCT-2002	(first entry)	DT	Db	266	VGE-ITGKPKMLPLPKMSLGPANFENDLNEAKBLKHVDTYRAKNIPDGYAIDFWRKKGGEN	324
DE	Alpha-isomaltosylglucosaccharide synthase #2.		XX	Oy	381	STAGLGENSHPYVVAQLAHNSRNQNYANGIQVHGGGGGIVTLDSTLGNFESHEV----436	
KW	Alpha-isomaltosylglucosaccharide synthase; sugar production;		KW	Db	325	NYGEFAWTANFEPSPRSATALKSND-ARGKIMG-----ITKPRIAKTDFFNNPTVQG	376
KW	cyclic tetrasaccharide; cyclo(-6)-alpha-D-glucopyranosyl(1-3);		KW	Oy	437	-----GHNY-GLHYVGDGFKGSYVHSAENNN--STVNGWDGDKKRF--IPNFYPSOTN	483
KW	alpha-isomaltosyltransferase; viscosity modifier; sweetener;		KW	Db	377	TDAASGGYFYPGHSEYKDFFIPVYVRSIDPYNPNAARSFWNHNSKDAFDKGTVGWMDFTD	436
KW	heat stabiliser; hydrolysis stabiliser; colour stabiliser; foodstuff;		KW	Oy	484	EKSCLNNOCOQEPFDGKFGFDAMAGGSPPSAANRFT-----MTPNSAIIQ-	530
OS	cosmetic; drug composition.		OS	Db	437	AVS-----SGGASIWFGNFTGAMSQALYEGQRAYTSNAQRVWQT	476
Unidentified.			PA	Oy	531	-RFF-----ENKAYEDSRSSSTGFSKWNADP-----554	
XX	W0200255708-A1.		XX	Db	477	ARTFYPGAAQRYATTLSGDIQYQTKGERINWAAGMEOARAYMLSSINGQVKWMDTGG	536
PN			XX	Oy	555	-----QMEP--YEHTIDRAEQ-----569	
XX	PA (HAYB ) HAYASHIBARA SEIBUTSU KAGAKU.		DB	537	FNQQDSTTNPNPDLYARNMOSALTTPVFRYHGNNHOQRPWYGSTABEASNEALQJRY	596	
PD	18-JUL-2002.		XX	Oy	604	NIYIPTAS--ADNRGSIILTINHEAGYNSYLIFINGDEKV-VSGQYKKSFVSDGQFWKERDV	660
XX	09-JAN-2002; 2002WO-JP00052.		PF	Db	657	EYLPAGTWDYDNROTLLTGQKISY-----AVNPDTLTDPLFTKGALIPSQ-KYDQY	710
XX	12-JAN-2001; 2001JP-0005441.		PR	Oy	570	-----ITASVNELESSEKMAELMTPY--AVYKVHM--WN-GNW-----TR 603	
XX	DR WPI: 2002-520129/55.		PR	Db	597	SLIPMYAVERSAYENGVLVRPLMQEYPADANAKNYLDAWMFCDWLALAPVYEKQQTSK	636
XX	DR N-PSDB; ABK88158.		XX	Oy	661	656 VDTREARKPKEQFCGPVT--TLVGYDPEGTLOSSYIYPAMYGAYGFTYSDDSQNLSDNDCQ	718
XX	New alpha-isomaltosylglucosaccharide synthase of bacterial origin for		PT	Db	711	NIYIPTAS--ADNRGSIILTINHEAGYNSYLIFINGDEKV-VSGQYKKSFVSDGQFWKERDV	660
PT	the production of cyclic tetrasaccharide gum -		PT	Oy	661	657 EYLPAGTWDYDNROTLLTGQKISY-----AVNPDTLTDPLFTKGALIPSQ-KYDQY	710
XX	The invention describes novel microbial polypeptides having		XX	Db	711	VGQAPVQTYDVFNTAQSSFTYDDDG--SSSYNES--GAY-FKQLMTAQ-----757	
XX	alpha-isomaltosylglucosaccharide synthase activity. The proteins are		PS	Oy	719	LOVDIKEGQDRFLANHRAANTVMNKFHINVPTESTOPATLVNNKILDTKSLTPAPEG	778
PS	useful for producing a sugar (I) having at least three glucose units		PS	Db	758	-----DNGSIALSETLG-----AKTGTSPALQSIV-----KLHGAG	791
XX	connected by alpha-1,4 bonds except for alpha-1,6 at the non-reducing		CC	Oy	779	LTYTYNG-----QALPAKENEGCIVSVNSGRYCLPVGORSGYSLPDWLYGQEYVV	829
CC	end. The invention also describes a method for the production of the		CC	Db	792	-----ASVTNSGAALASYASLQALAKASASEG-----WAKGKD1YQ	826
CC	cyclic tetrasaccharide cyclo(-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-		CC	Oy	830	D-----SGAKA 835	
CC	glucopyranosyl(1-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-glucopyranosyl		CC	Db	827	DVTTYKLSAGAAA 839	
CC	(I)-(II) by treatment of (I) with alpha-isomaltose transferase.		CC				
CC	(II) and similar sugars in crystalline or syrup form are used as		CC				
CC	a viscosity modifier, sweetener, heat stabiliser, hydrolysis stabiliser		CC				
CC	and colour stabiliser for foodstuffs, cosmetics and drug compositions.		CC				
CC	This is the amino acid sequence of an alpha-1-isomaltosylglucosaccharide		CC				
XX	synthase.		SQ	Sequence	1286 AA;		
SQ						RESULT 13	
						AAB69136	
						ID AAB69136 standard; Protein: 2314 AA.	
						XX AC AAB69136;	
						XX AC AAB69136;	

DT	24 - APR - 2001	(first entry)
XX	M. catarrhalis les1	200kDa protein SEQ ID NO:1.
DE		
XX	Moraxella catarrhalis strain Q8; major outer membrane protein;	
XX	200kDa outer membrane protein; antibacterial; immunogenic; infection;	
KW	otitis media; detection.	
XX	Moraxella catarrhalis.	
OS		
XX	PN	WO200107619-A1.
XX	XX	
PD	01-FEB-2001.	
XX	XX	
PF	26-JUL-2000; 2000WO-CA00870.	
XX	XX	
PR	27-JUL-1999; 99US-0361619.	
XX	XX	
PA	(CONN)- CONNAUGHT LAB LTD.	
XX	XX	
PI	Loosmore SM, Sasaki K, Yang Y, Klein MH;	
XX	XX	
DR	WPI: 2001-159722/16.	
DR	N-PSB; AAF59105.	
XX	XX	
PT	New nucleic acid encoding Moraxella catarrhalis outer membrane protein,	
PT	useful in protective vaccines and for diagnosis -	
XX	XX	
PS	Claim 1: Fig 5A-Y; 247pp; English.	
XX	XX	
CC	The present invention describes an isolated and purified nucleic acid (I)	
CC	which encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.	
CC	The 200 kDa outer membrane protein (II) has antibacterial activity and	
CC	can be used in vaccines. (II), and its truncated versions, are used as	
CC	immunogenic compositions and vaccines to protect against M. catarrhalis	
CC	infections, particularly otitis media in humans. (III) is also used as	
CC	antigen-in immunoassays for detecting specific antibodies (Ab), and to	
CC	generate Ab. (I) are used for recombinant production of (II) and its	
CC	fragments are used as probes for identifying/cloning 200 kDa protein	
CC	genes from other strains, and for diagnostic detection of M. catarrhalis.	
CC	(I) makes possible production of large amount of recombinant immunogens.	
CC	Exposure of truncated versions of (II) reduces toxicity of the protein	
CC	towards the Escherichia coli host. The present sequence represents the	
CC	M. catarrhalis les1 200kDa protein, which is given in the exemplification	
CC	of the present invention.	
XX	Sequence 2314 AA:	
Qy	Query Match 3.1%; Score 144; DB 22; Length 2314;	
	Best Local Similarity 20.1%; Pred. No. 0.18;	
	Matches 212; Conservative 129; Mismatches 388; Indels 326; Gaps 52;	
Db	9 TILAPLAIGYESATA----ADNNSAIYFNTSQINLDQSLAEVKAQSQQLPAHKEP 63	
Db	47 TRVATIALIVGATLNGSAYAQNNNSKIAFCTTG--NDNDASASIEASTAIGSLAKAHANQ 104	
Qy	64 ----GDSOP-----HLTSLRKSLLYLVPVKADDKTPVQEARD--DNNKIL 103	
Db	105 AIAIGSSKPDPRNQANOKAGSHAKG-KESTAIGGDLAEGDASTAIGSDDLYLDRN-- 160	
Qy	104 GTLTLPPSSLPTDHWLDGYPEGGIIDFTPHNGTKKIINTVAEVNKLSDASGSIIHS--H 161	
Db	161 STNSKPGNLSTLQNHNTVLRQ--TIDSNSQKVRFAAE-GHASTAVGAMAYAKHG 215	
Qy	162 LTNNALVEITANTGRWYRDITYLPOGFDLEGMRVRESSA--GYSSTTFYGDKRKVTLSVG 218	
Db	216 FA-NAFTRSTRATEGNISLAVGLTAKAE-KGTVIAIGSNAOAINGALAAGDPRVDLTDY 273	
Qy	219 NTLLF-----KVV--NGQWFRSGELENNRRTYAQHWSAELPAHWWPGL 261	
Db	274 IALGYSQIQLNNNNNNKATVYPEGGNSNIKSSKATGNGLF---SIGSSTIKRINNTYGA 330	
Qy	262 -----NLVIKGNNISGRNLNDI---KIG-----APGEILLHTID 291	

Db	331	GYEDDAVNQLQKAVENLAKRQITFGDDNGTGVVKVALGETLTIKGGETQADKLTDNNN	390
OY	292	IGMLTTPRD--RFDFADKEAHEAYFOTIPVSMAV-----NIVAPLHLKEV-MLP	335
Db	391	IGVVTDTNTGLERYKLAKLNSG---LEPVSTKLTASERKVTVGSNTAELLSGCCSLLTFTP	446
QY	340	TCELLTID-----MDPGNGGMHESGTMRORIGKELYLVSHGIDNARYGLNSTAGLG	388
Db	447	TINASIDTKTVGTDGLKETDNTSNTALEDT--RITDKI-----GFSNKAGTYDE	494
OY	389	SHPYY-NYAQLAHNNSRGNYANGIQYHGSSGGGIVYTJLQSTLGNFESHEVGHNTGLGHYD	447
QY	495	NRPYLDKDLKVGNSTLN-----NNG--LTWNNTIG-----GSNRKOIQVGAD	534
Db	448	GFKGSVHSRFAEENNNSTRGR-----DG-----DKKRF-----473	
OY	535	GTFKADLVNVSIAAEKFQTRTIEEFIGADAKVDKSPYLDKKQLOQVGGYKITKDSC	594
Db	474	-IPNPFQPSQTNEKRSCLN---NOCQEPFPDGHFKFGFDAM-AGCSPPSAAANRFTM-	520
OY	595	INAGDQKLSN-VKTDATDDDAVITYKQLKQVQDAGALOSFSTIDEKGQEFETISNLYSNG	653
QY	521	YTPNNSAIIQREFENKAVEDDSRSSTGSKWNADTQEMEPYEHTIDRAEQTTASYNELSES	580
Db	654	NTPNTFETITFAGENGISIISNDIAKGVKVRGID-----PINGLTTP	694
QY	581	KM---AELMAEYAVKVKEMWANGNTRNI---YIPT-ASADNRGSLT-----620	
Db	695	KIYGYSQDKGKTQVIEYAVASGNTKNIIRGLSPPLSPSLTNAGGTYRTTEQGNTITSDEK	754
OY	621	-----INHEAGNSYLFINGDEKVSQCY---KKSFVSDGQF	654
Db	755	SKAASIGDILNTGFNLKNNNSNSVGFSTNTVDFDGNATTAKTYDEINOTSKVYDYN	814
OY	655	WKERDVDTREARKPEQFGPYVTTLYGYDPEGTLSYIYPAMYTAGGYFTSYDDSONLSD	714
Db	915	VDERTIELTGDNGTINRKVKTTLT-----TINA-----NGKATNFNST	853
QY	715	NDCOLQVDTKEGQRLFRFLANHRANTYMMKHFHINYPTESOPTAQLVCNNKILDTKSLTP	774
Db	854	TNDDALVNAKD----IEENL--NLTAKEIHTTKGTAIDLQTFKVKKGATDDEITY	905
QY	775	APEGLTYYNGQ--ALPAKENEGCILIVSYNSGKRYCLPVQRSGYSLPDWIVQEVYVD	831
Db	906	GKDG--TONGKTYNTLKLKGENGITVATNK-----DGTVTFGINTQS	945
QY	832	GAKARVLLSEMDWNLNSYRNRIGEFYGNVNPADMKKV	866
Db	946	GLKA-----GDSITLNKDG--LSTKPNASNEQIO	972
RESULT 14			
ID	ABG30538	standard: Protein: 1251 AA.	
XX	ABG30538;		
AC			
XX			
DT	07-OCT-2002	(first entry)	
XX			
DE		Alpha-isomaltosylglucosaccharide synthase # 2 mature protein.	
XX			
KW		Alpha-isomaltosylglucosaccharide synthase; sugar production;	
KW		cyclic tetrasaccharide; cyclo(-6)-alpha-D-glucopyranosyl(1-3);	
KW		alpha-isomaltosyl transferase; viscosity modifier; sweetener;	
KW		heat stabiliser; hydrolysis stabiliser; colour stabiliser; foodstuff;	
KW		cosmetic; drug composition; enzyme.	
XX			
OS		Unidentified.	
XX			
PN		WO200255708-A1.	
PD		18-JUL-2002.	



CC probiotic specimens, probiotic-containing food, feed and fodder, by  
 CC introducing the plasmidic gene into the structure of the bacterial  
 CC strain, preferably L.lactis IBB501, IBB502, IBB140 and  
 CC subsequently using the obtained bacterial strain in plant material and  
 CC milk fermentations or in biomass production. The present sequence  
 CC represents the lactococcus amylolytic enzyme.

xx Sequence 784 AA:

Query	Match 3.0%	Score 142;	DB 24;	Length 784;
Best Local Matches 141;	Similarity 20.0%;	Pred. No. 0.046;		
Matches 141;	Conservative 95;	Mismatches 283;	Indels 186;	Gaps 36;
Db 239	MRITYAQHITMSELPAHWIYVPLGLNVIROGSNLNIGRLLKAPGELLHTADIGMLRTTP	298		
Db 122	NHATNTIGVLTIVLTVPGN--QNGMTAYAKLTAEQTVDYAGSTYGTLSSTS--NTT	175		
Qy 299	RDRDFDFAKDEAHREYFOTIPVSRMIVNNYAPHLKEYMVLPTGELLTMDPGNGGWHSGT	358		
Db 176	NDPYSIATTGGNRSVVES--SANLASNL-----LAQKSAT-----WRYAS	216		
Qy 359	MROIGKELYSHGIDNANTGLNSTAGLGENSEHVNGLCHYVD--GFKGSYHRSAAENNNTWGNGDKRPI	418		
Db 217	PTQATVDEL-HIRD---TSSSTSGV-----SAGRKGKELGVQSQTDDPN	258		
Qy 419	GIVITLDSTLGEFNSHEVGHNYGLCHYVD--GFKGSYHRSAAENNNTWGNGDKRPI	474		
Db 259	TGTAT----GLDYLKNEGTNYNIQLHASCVQVQSGNRTTAQPNNNWGDQPMKWWQ	313		
Db 475	PNFYPQSQNEKSCLANNOQEPFDG-HKFGFDAMAGGSPSAARFTMPNNSAIIORFF	533		
Qy 314	SEYASISVSNEVTRI-NEMKEMVOGLHTNGTSVYRD---MVLNH--VYSQSASAFERA--	364		
Qy 534	ENKAYFDSRSTG-----FSKWNADTOQEPEYTIID-RABOTASYNELES	580		
Db 365	EPGYFPRKNTSGCGNDTASNHEAFGKYIIDSVTYWAQNYDIDGFRDEMTL---LDST	420		
Qy 581	KHAELMAEYAVVKHM-----WNGNWTRNVIYPTASADNRGSLTLINHEAGYNSYLFIN	634		
Db 421	TWNKLRAALTAIDPHITMYGGWGSNANN-1PETSINYKNPFIG-----FFNP	470		
Qy 635	GDEKVYSQ--CYKKSFVSDGOFWKERDVDTREARKPEQFGVPVTTLV-----GYD	684		
Db 471	GERDAISNNCGSAGGFAA-----GNTASTIVAGALLASGWN	508		
Qy 685	PEGTLSYYKIPAKGAY----GFTYSDSDONLSNDNCOLOVTKREGQFLRLANHANN	739		
Db 509	GNGTVQAFLTTSQSINTVECHDSFLNDLSWADEND-----SYATHQARY	554		
Qy 740	TYPAHKHI--NPVTSQOPTAQLVNNKILDKSLTPAPGLT-YTNGQALPA-----	790		
Db 555	TIANATNLLANGTIVMETGE--PEQSKLYNPSNITPLSPTQDQYQSGSMERAYRAS	612		
Qy 791	-----KENEGCIVSYNSGRKRYCLPVGQRSSYSLPDWIVQEYVDGAKVTLISDMDNL	845		
Db 613	WTATAKNSYNGLGLSNGTY-----GNYWP---GSNLVTPVAGDVNAMWDNV	660		
Qy 846	SYNRIG-EFGNNPADMKKKKARNQOYL--DFSK---PRSMRY	883		
Db 661	MDNQNAVNFIQN----LMRFKKSNPQFWDDYSKLAWTPSIGV	700		

Search completed: October 8, 2003, 16:24:44  
 Job time : 59.919 secs

GenCore version 5.1.6  
Copyright (C) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw mode1

Run on: October 8, 2003, 16:19:57 ; Search time 47.5173 Seconds

(Without alignments)  
4811.606 Million cell updates/sec

Title: US-10-002-309B-2

Perfect score: 4709 Gapext: 0.5

Scoring table: BLOSSUM62

Sequence: 1 MKLKVLSCTTIAPLAIGVFS..... AWNGQYLDPSKPRSMRVMVYK 886

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBLE23:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rabbit:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_virus:\*

16: sp\_bacteriophage:\*

17: sp\_archaea:\*

SPTREMBLE23:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rabbit:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_virus:\*

16: sp\_bacteriophage:\*

17: sp\_archaea:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rabbit:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_virus:\*

16: sp\_bacteriophage:\*

17: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the total score distribution, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4709	100.0	898	2 Q9ZAL1	082882 escherichia
2	154.8	32.9	310	2 Q9AGX6	Q9zal1 escherichia
3	686.5	14.6	1002	2 Q9AGX6	Q9agx6 vibrio chol
4	671	14.2	1335	16 Q9KN18	Q9kn18 vibrio chol
5	571	12.1	875	16 Q8D5P4	Q8d5p4 vibrio vuln
6	290	6.2	957	16 Q9KTH2	Q9kth2 vibrio chol
7	166.5	3.5	1217	16 Q9GW3	Q9gw3 clostridium
8	161	3.4	1038	2 Q45780	Q45780 bacteroides
9	159.5	3.4	2367	2 Q6034	Q46034 clostridium
10	158.1	3.4	2204	2 Q6155	Q81170 plasmid
11	156.5	3.3	1157	16 Q9TF62	Q9tf62 clostridium
12	154.5	3.3	2367	2 Q9F931	Q9f931 clostridium
13	150	3.2	3692	2 Q8RR3	Q8kr3 fusobacteri
14	149	3.2	1284	2 Q8RQU9	Q8rq9 bacillus
15	148.5	3.2	2468	16 Q912M3	Q9l1m3 pseudomonas
16	148	3.1	1021	2 Q46085	Q46085 clostridium

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the total score distribution, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	148	3.1	2937	16 Q8D990	vibrio vuln
2	147.5	3.1	1530	16 Q8Y479	listeria mo
3	147.5	3.1	1093	16 Q8F107	leptospira
4	147	3.1	1377	16 Q8EV69	mycoplasma
5	147	3.1	1386	16 Q9Bz78	mycoplasma
6	146.5	3.1	1196	17 Q92DL0	listeria in
7	146.5	3.1	2057	2 Q9TS48	methanosaarc
8	145.5	3.1	2189	5 Q9RE05	leuconostoc
9	145.5	3.0	1395	2 Q9AISO	Q9aiso staphylococ
10	143.5	3.0	1284	16 Q8ev71	mycoplasma
11	143	3.0	1110	17 QBPXT0	Q9Pxd1 uncultured
12	140	3.0	1135	1 Q9579	Q8XAN9 escherichia
13	139	3.0	1270	16 Q8XAN9	Q9g13 streptococc
14	138.5	2.9	1200	17 Q9LCH3	Q9t114 methanosaarc
15	138.5	2.9	1575	2 Q82KG6	Q9f114 methanosaarc
16	138	2.9	1964	2 Q8TL14	Q9kmg9 moraxella c
17	138	2.9	2675	5 Q9f285	Q813F9 plasmid
18	138	2.9	2894	17 Q58791	Q5B791 methanococc
19	137	2.9	1480	2 Q8CC13	Q95579 helicobacte
20	137	2.9	5559	16 Q82KG6	Q9zg96 salmonella
21	137	2.9	1457	2 Q9f289	Q9f289 yersinia pe
22	136.5	2.9	3705	2 Q9f285	Q9f285 yersinia pe
23	136.5	2.9	3705	16 Q8ZHA1	Q8zha1 yersinia pe
24	136.5	2.9	3710	16 QBCZU2	Q8cu2 yersinia pe
25	136	2.9	1677	16 Q8RGN7	Q8rgn7 fusobacteri
26	136	2.9	1980	3 Q05747	Q05747 saccharomy
27	136	2.9	2167	16 Q92EK5	Q92ek5 listeria in

## ALIGNMENTS

RESULT :	ID	PRELIMINARY;	PRT;	898 AA.
1	OB2882	OB2882; Q9ZGU1;	AC	01-NOV-1998 (TREMBLrel. 08, Created)
			DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)
			DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)
			DE	Plasmid PO157 DNA, complete sequence.
			GN	L7031.
			OS	Escherichia coli O157:H7.
			OX	Plasmid PO157.
			OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
			OX	NCBI_TAXID=83334;
			RN	[1]
			RP	SEQUENCE FROM N. A.
			RC	STRAIN=O157:H7;
			RX	MEDLINE=98290540; PubMed=9628576;
			RA	Makino K., Ishii K., Yasunaga T., Hattori M., Yokoyama K., Yutsudo H.C., Kubota Y., Yamaichi T., Iida T., Yamamoto K., Honda T., Han C.G., Ohtsubo E., Kasamatsu M., Hayashi T., Kuhara S., Shinagawa H.;
			RA	"Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic Escherichia coli O157:H7 derived from Sakai outbreak." [2]
			RA	"Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic Escherichia coli O157:H7 derived from Sakai outbreak." [2]
			RL	DNA Res. 5:1-9(1998).
			RL	[1]
			RP	SEQUENCE FROM N. A.
			RC	STRAIN=BDL933;
			RX	Burland V., Shao Y., Perna N.T., Plunkett G., Sofia H.J., Blattner F.R.;
			RA	"The complete DNA sequence and analysis of the large virulence plasmid of Escherichia coli O157:H7." [3]
			RT	Nucleic Acids Res. 26:4196-4204(1998). [3]
			RL	[1]
			RP	SEQUENCE FROM N. A.
			RC	STRAIN=BDL933;
			RA	Brundier W.;

Submitted (MAR-1997) to the EMBL/GenBank/DDJB databases.

DR EMBL; AB011549; BAA31157.3; -.

DR EMBL; Y11831; AAC70099.1; -.

KW Hypothetical protein; Plasmid.

SQ SEQUENCE 898 AA: 9548-9548;

Query Match 100.0%; Score 4709; DB 2; Length 898; Best Local Similarity 100.0%; Pred. No. 1.4e-40; Matches 886; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLYKLSCTILAPLAGTSAVTCATTAADNSAIVNTSOPINDLGSLAAEVKAQSQILPAH 60

Db 13 MKLYKLSCTILAPLAGTSAVTCATTAADNSAIVNTSOPINDLGSLAAEVKAQSQILPAH 72

Qy 61 PKEGSQPHLTSRKSLVLLVPRYKADDKTPVQEARDNNKLGTLTLYPSSLPDTIH 120

Db 73 PKEGSQPHLTSRKSLVLLVPRYKADDKTPVQEARDNNKLGTLTLYPSSLPDTIH 132

Qy 121 LDGVEGGIDFDPHTNGTKRINTVAYENKLDAASGSSTHSILTNNALEIHTANGRWRD 180

Db 133 LDGVEGGIDFDPHTNGTKRINTVAYENKLDAASGSSTHSILTNNALEIHTANGRWRD 192

Qy 181 IYLPGPDLEGKAVREVSSAGSSTVFGDRKVTLSGNTLFLKYYNGQWFFSGELENRR 240

Db 193 IYLPGPDLEGKAVREVSSAGSSTVFGDRKVTLSGNTLFLKYYNGQWFFSGELENRR 252

Qy 241 ITYAOHWAELPAHWHWVGLNLYIKQGNLSGRNDIKIGAPELLHTIDGMLTPRD 300

Db 253 ITYAOHWAELPAHWHWVGLNLYIKQGNLSGRNDIKIGAPELLHTIDGMLTPRD 312

Qy 301 RFDFAKDEKAHREYFOTIPVSRYMNTYNNAPHLKVEMLPTGEILITMDPGNGWHSQTMR 360

Db 313 RFDFAKDEKAHREYFOTIPVSRYMNTYNNAPHLKVEMLPTGEILITMDPGNGWHSQTMR 372

Qy 361 ORIGELVSHGIDDNANYGLNSTAGLGENSEHPVVAQLAHNSRGNYANGIQVHGGSGGG 420

Db 373 ORIGELVSHGIDDNANYGLNSTAGLGENSEHPVVAQLAHNSRGNYANGIQVHGGSGGG 432

Qy 421 IYTLDSTLGNFSHEVGHANGLGHVDFGPKVERSAAENNSTMWGDDKKRKFIPNYP 480

Db 433 IYTLDSTLGNFSHEVGHANGLGHVDFGPKVERSAAENNSTMWGDDKKRKFIPNYP 492

Qy 481 QTENEKSLCLANICQCEPDGHKRGFDAMAGGSPFSAANRFTMTPNSATIQRFENKAVID 540

Db 493 QTENEKSLCLANICQCEPDGHKRGFDAMAGGSPFSAANRFTMTPNSATIQRFENKAVID 552

Qy 541 SRSSGFSKWADTOEMEPYHTIDRAEQTASTNELSEPSKMAELMAEAVVHMGNGN 600

Db 553 SRSSTOFSKWADTOEMEPYHTIDRAEQTASTNELSEPSKMAELMAEAVVHMGNGN 612

Qy 601 WTRNITYLPTASADNRCSILTINHEAGYNSYLINGDEKYSQGKSPVSDGFWKRDV 660

Db 613 WTRNITYLPTASADNRCSILTINHEAGYNSYLINGDEKYSQGKSPVSDGFWKRDV 672

Qy 661 VDTREARKPEFGPVPTLVGYDPEGTSSYTYIPMAGAYGFTYSDSONLSDNDCQLQ 720

Db 673 VDTREARKPEFGPVPTLVGYDPEGTSSYTYIPMAGAYGFTYSDSONLSDNDCQLQ 732

Qy 721 VDTKEGOLRFLRFLHANRNTMANKFHINVPTEESOPTAQATLVCNNKILDTSLPAPEGLT 780

Db 733 VDTKEGOLRFLRFLHANRNTMANKFHINVPTEESOPTAQATLVCNNKILDTSLPAPEGLT 792

Qy 781 YTNGQQLAKENEGCLVSYNSGKRYCILPVGQRSGYSLDPMVQEVYDGSAKAKVLLS 840

Db 793 YTNGQQLAKENEGCLVSYNSGKRYCILPVGQRSGYSLDPMVQEVYDGSAKAKVLLS 852

Qy 841 DWDNLSNRIGEFVGVNNPDKKVKAWNGCYLDSSKPMSRMVYK 886

Db 853 DWDNLSNRIGEFVGVNNPDKKVKAWNGCYLDSSKPMSRMVYK 898

O9ZAL1 PRELIMINARY; PRT: 310 AA.

ID O9ZAL1  
AC O9ZAL1  
DT 01-MAY-1999 (TREMBrel. 10, Created)  
DT 01-MAY-1999 (TREMBrel. 10, Last sequence update)  
DT 01-DEC-2001 (TREMBrel. 19, Last annotation update)  
DE Hypothetical 34.0 kDa protein (Fragment).  
OS Escherichia coli.  
NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC Brunder W.; Submitter (FEB-1997) to the EMBL/GenBank/DDJB databases.  
DR EMBL; Y11275; CAAT2142.1; -  
KW Hypothetical protein; Plasmid.  
FT 310  
SQ SEQUENCE 310 AA: 34032 MP: 2AA72C166F315BA6 CRC64:  
1 MKLYKLSCTILAPLAGTSAVTCATTAADNSAIVNTSOPINDLGSLAAEVKAQSQILPAH 60  
Query Match 32.9%; Score 1548; DB 2; Length 310;  
Best Local Similarity 100.0%; Pred. No. 1.2e-33;  
Matches 298; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 133 MKLYKLSCTILAPLAGTSAVTCATTAADNSAIVNTSOPINDLGSLAAEVKAQSQILPAH 72  
Db 133 MKLYKLSCTILAPLAGTSAVTCATTAADNSAIVNTSOPINDLGSLAAEVKAQSQILPAH 72  
Qy 61 PKEGSQPHLTSRKSLVLLVPRYKADDKTPVQEARDNNKLGTLTLYPSSLPDTIH 120  
Db 61 PKEGSQPHLTSRKSLVLLVPRYKADDKTPVQEARDNNKLGTLTLYPSSLPDTIH 132  
Qy 723 PKEGSQPHLTSRKSLVLLVPRYKADDKTPVQEARDNNKLGTLTLYPSSLPDTIH 132  
Db 723 PKEGSQPHLTSRKSLVLLVPRYKADDKTPVQEARDNNKLGTLTLYPSSLPDTIH 132  
Qy 120 IYLPGPDLEGKAVREVSSAGSSTVFGDRKVTLSGNTLFLKYYNGQWFFSGELENRR 180  
Db 120 IYLPGPDLEGKAVREVSSAGSSTVFGDRKVTLSGNTLFLKYYNGQWFFSGELENRR 192  
Qy 121 LDGVEGGIDFDPHTNGTKRINTVAYENKLDAASGSSTHSILTNNALEIHTANGRWRD 180  
Db 121 LDGVEGGIDFDPHTNGTKRINTVAYENKLDAASGSSTHSILTNNALEIHTANGRWRD 192  
Qy 122 IYLPGPDLEGKAVREVSSAGSSTVFGDRKVTLSGNTLFLKYYNGQWFFSGELENRR 180  
Db 122 IYLPGPDLEGKAVREVSSAGSSTVFGDRKVTLSGNTLFLKYYNGQWFFSGELENRR 192  
Qy 123 LDGVEGGIDFDPHTNGTKRINTVAYENKLDAASGSSTHSILTNNALEIHTANGRWRD 180  
Db 123 LDGVEGGIDFDPHTNGTKRINTVAYENKLDAASGSSTHSILTNNALEIHTANGRWRD 192  
Qy 181 IYLPGPDLEGKAVREVSSAGSSTVFGDRKVTLSGNTLFLKYYNGQWFFSGELENRR 240  
Db 181 IYLPGPDLEGKAVREVSSAGSSTVFGDRKVTLSGNTLFLKYYNGQWFFSGELENRR 252  
Qy 182 IYLPGPDLEGKAVREVSSAGSSTVFGDRKVTLSGNTLFLKYYNGQWFFSGELENRR 240  
Db 182 IYLPGPDLEGKAVREVSSAGSSTVFGDRKVTLSGNTLFLKYYNGQWFFSGELENRR 252  
Qy 183 IYLPGPDLEGKAVREVSSAGSSTVFGDRKVTLSGNTLFLKYYNGQWFFSGELENRR 240  
Db 183 IYLPGPDLEGKAVREVSSAGSSTVFGDRKVTLSGNTLFLKYYNGQWFFSGELENRR 252  
Qy 241 ITYAOHWAELPAHWHWVGLNLYIKQGNLSGRNDIKIGAPELLHTIDGMLTPRD 300  
Db 241 ITYAOHWAELPAHWHWVGLNLYIKQGNLSGRNDIKIGAPELLHTIDGMLTPRD 312  
Qy 253 ITYAOHWAELPAHWHWVGLNLYIKQGNLSGRNDIKIGAPELLHTIDGMLTPRD 312  
Db 253 ITYAOHWAELPAHWHWVGLNLYIKQGNLSGRNDIKIGAPELLHTIDGMLTPRD 312  
Qy 301 RFDFAKDEKAHREYFOTIPVSRYMNTYNNAPHLKVEMLPTGEILITMDPGNGWHSQTMR 360  
Db 301 RFDFAKDEKAHREYFOTIPVSRYMNTYNNAPHLKVEMLPTGEILITMDPGNGWHSQTMR 372  
Qy 361 ORIGELVSHGIDDNANYGLNSTAGLGENSEHPVVAQLAHNSRGNYANGIQVHGGSGGG 420  
Db 361 ORIGELVSHGIDDNANYGLNSTAGLGENSEHPVVAQLAHNSRGNYANGIQVHGGSGGG 432  
Qy 421 IYTLDSTLGNFSHEVGHANGLGHVDFGPKVERSAAENNSTMWGDDKKRKFIPNYP 480  
Db 421 IYTLDSTLGNFSHEVGHANGLGHVDFGPKVERSAAENNSTMWGDDKKRKFIPNYP 492  
Qy 433 IYTLDSTLGNFSHEVGHANGLGHVDFGPKVERSAAENNSTMWGDDKKRKFIPNYP 492  
Db 433 IYTLDSTLGNFSHEVGHANGLGHVDFGPKVERSAAENNSTMWGDDKKRKFIPNYP 500  
Qy 481 QTENEKSLCLANICQCEPDGHKRGFDAMAGGSPFSAANRFTMTPNSATIQRFENKAVID 540  
Db 481 QTENEKSLCLANICQCEPDGHKRGFDAMAGGSPFSAANRFTMTPNSATIQRFENKAVID 552  
Qy 541 SRSSGFSKWADTOEMEPYHTIDRAEQTASTNELSEPSKMAELMAEAVVHMGNGN 600  
Db 541 SRSSGFSKWADTOEMEPYHTIDRAEQTASTNELSEPSKMAELMAEAVVHMGNGN 612  
Qy 601 WTRNITYLPTASADNRCSILTINHEAGYNSYLINGDEKYSQGKSPVSDGFWKRDV 660  
Db 601 WTRNITYLPTASADNRCSILTINHEAGYNSYLINGDEKYSQGKSPVSDGFWKRDV 672  
Qy 661 VDTREARKPEFGPVPTLVGYDPEGTSSYTYIPMAGAYGFTYSDSONLSDNDCQLQ 720  
Db 661 VDTREARKPEFGPVPTLVGYDPEGTSSYTYIPMAGAYGFTYSDSONLSDNDCQLQ 732  
Qy 721 VDTKEGOLRFLRFLHANRNTMANKFHINVPTEESOPTAQATLVCNNKILDTSLPAPEGLT 780  
Db 721 VDTKEGOLRFLRFLHANRNTMANKFHINVPTEESOPTAQATLVCNNKILDTSLPAPEGLT 792  
Qy 781 YTNGQQLAKENEGCLVSYNSGKRYCILPVGQRSGYSLDPMVQEVYDGSAKAKVLLS 840  
Db 781 YTNGQQLAKENEGCLVSYNSGKRYCILPVGQRSGYSLDPMVQEVYDGSAKAKVLLS 852  
Qy 841 DWDNLSNRIGEFVGVNNPDKKVKAWNGCYLDSSKPMSRMVYK 886  
Db 841 DWDNLSNRIGEFVGVNNPDKKVKAWNGCYLDSSKPMSRMVYK 898  
Qy 853 DWDNLSNRIGEFVGVNNPDKKVKAWNGCYLDSSKPMSRMVYK 900  
Db 853 DWDNLSNRIGEFVGVNNPDKKVKAWNGCYLDSSKPMSRMVYK 908

Hydrolase	SEQUENCE	1002 AA;	114645 MW;	6A239DEB6408407F CRC64;	PRT;	1335 AA.
Query Match	Score 686.5; DB 2; Length 1002;					
Besti Local Similarity 24.0%; Pred. No. 2.8e-36;	Mismatches 157;	Indels 295;	Gaps 40;			
Matches 230;	Conservative 157;	Mismatches 230;	Indels 295;	Gaps 40;		
27 NSAIYFNITSQINDLQGLSAAEYFAQSQIPLPAHPKEGDSOPHILSTSRLSLLVLP-VKA 85						
137 NEIDYIHTTEIEKVNGLSLEGEREVQTHV--SPEGRKNEPPEITIGDADLIFKPSIKN 194						
86 DDKTPQYQEARDNNKKLQLTLVPPSSLQPTDTYHLDGYPEGGSDFTPHNGTKIINTVA 145						
195 SSISLMKIKYSEDG--LTSKVYMASPSMPLPKTIDOPID-----ID----- 230						
146 EVNKLDASGSIIHSHTLNNALEIHTANGRWRDYLFLQGPDELEGKAVRFFVSAGYSST 205						
231 --	--	--	--	--	--	--
206 VFGDRKVTLSVGNNTLFLKYVNGOMFERSGLENNR-ITYAQHTRWASAEELPARWLVPGMLNV 264						
231 --	--	--	--	--	--	--
ENKKVVSYNSNSTKSAAEIPWNKOKSGSILH 25.9						
265 IKQGNLSGRL---NDIKIGAPGELLHLTIDIGMLTTPRDFDAKEAHR-EYFOTT 31.8						
260 FEDEN -GNLGIIESERIKFSAPS ELLIONIDGLYKPGRNIVVIKELTAYDYFOKV 31.7						
319 PVSRMIVNNYATPLAHKEYMMPGTLGEELLTDMDPGKGWHSGSTMQRQIGKELYSHGTIDNANYG 37.8						
318 PVSKLFLSDTYPTKPIRKITLPGSYTEKSADIGWHOGDRAEVGKALVSTGINNANG 37.7						
379 INSTAGLGENSHPYVVAQLAHHNSRGNYANGIQYHGGSGGGIVTLDSTLGNESFSEYGH 4.38						
378 IVASSGYSQ -QYNRLTNHITAHTNIGYNNGVVHGGSGGGIVTLENTLNHNWSHELGH 4.36						
439 NTRGLGHYVTDGFKGSVIRSAENNNSTGWDGDKKRIPNPFYPSQNEKSCS - MNOCQEPF 4.96						
437 NYGLGHYAG -GTSH ---GPDTNSWGWDGYKTRTANFDWKRSPOSNTRDPDQEYVVPF 4.90						
497 -DGHKEFDAMGGSPP - SANRRTMTPNSAIIORFENKAVIDFSRSRSTSFSKWNAD 5.53						
491 MOKTYLWDASSGGDHQNGTISRYTLLHHYVARIQDWLKGAVV--INNDMMWD-E 5.46						
554 TOEMEPYHTIDRAEQITASYNLESESKMAELMAEAYVVKHMANGNTRNIYIPTASAD 6.13						
547 Liniyyikrgt----NFKPIKKGVPVTTLGVDPKIN----PSQLYPPPTS-- 5.91						
614 NRGSLI-----TINHEAGNSY--LFINGDEKVVSQ---- 6.42						
592 NYGNIFDLEKPSESSLKGROYKDNYLDVNTNHMTLYNKRKEKICFSYSLPKGKK 6.51						
643 ---GYKK-----SEVSDSOOFWKEDDVDTREARKPEOFGVPTTILV--Y 6.82						
652 FEFLGYEDIENKICTGSRSRTHYLEDG --KKNPI---ESKNYDF--LTSIDSGDETS 7.01						
683 YDPEGTL-----SSTYIYPAMYGA-----YGFYKSDSOSNLS--- 7.13						
702 YPDSTIGESKCSLMSMGTVYAGFLKGNSCQRIDGVFMNGFQRAFTLHQSGYSTYIW 7.61						
714 DNDQLODVTKEQSLR-FRLANHRNNTVMKFRHINYVPTESQPTQATLVCNNKILDKTSK 7.72						
762 SNCVCLXKTKDKNMIES1SPVYRIEKENOSNRHLNISREPPIIDINVYCGHEHLSIKV 8.21						
773 TPAEPLGTTVNGQALPAKENEGCIVSYNSGKRYCLPVGQRSGYSLPDWIVQGYVYDSG 8.32						
822 SDNPD-----IKLRLGPIL-----VGOEHGYT-----SYEPKLPSG 8.52						
833 ARKVLSSWDNIDS-YRIGETPVGNPADMINKVRWANGQYL--DFS-KPRSMRVV 8.84						
853 ----WFKHNDNFPKNEINHGLKMRVND----NDEYICRFENFSDDREMKFV 8.97						
549 KWNADTOEMEPYHTIDRAEQI----- 10.1						
595 YNRFNHN-----ITAHTNNGVYTKKDDLPVQVHGGSGGIVTLEATGNGNESHLDG----- 10.1						
438 HNTGLGHYDGFKGFSVHRSAAENNNSTGWDGDGDKKRETPNF -----YPSQTNEKSCLNNC 10.1						
649 HNTGLGHW-----PYMASIH-----DLESCHGWDAFHQRFPGNLHWKGDDVYTOOGD----- 10.1						
492 CQEFP-DGKEFDPAMAGGSP - FSAANRFTMAYTPNSAIIORFFENKAVIDFSRSSTGF 10.1						
697 IVPPFKDAFRELDAQNGEQTYVTG15871SPKTFLEHQAOSRKQAORMNGFNLDSHSPSGK 10.1						
549 KWNADTOEMEPYHTIDRAEQI----- 10.1						

Db	757 QWDQETQRYKAVETDTPKPQQGVPPVTLGGTYDPQNENPSQIYPLVSNYGNFELPPQ 816	229 DVYTQQQGD-----DIVPFKDAFREFLVDAQNGGEQYVGTSIRFTLEPAQSREAQRNW 282
Qy	581 KHAELMAEAYVKVHAWNNGNTRNIYITMASADNRGSLTINRHEAGNSY--LFLINGDEK 638	533 FENKAVFTDSSTCFSKNAHDTOBEPHTIDAEQTASFNEL-----SFSKMA 58:
Db	817 EQEKEYOLE-----GW-----QAAGD-----LT-QAEIQYNOWQTLTIDGQQL 852	283 MNNGFNLDSSSPSGYVQNTQAAQRQYOTVEDTPKPOQTGVAATLGLYDPPYNESQI - 341
Qy	639 VV-----SQGYKSFVSDGQFWKERDYVD-TREARKPEQFGPVPTTLYGYDPEGTLS 690	584 ELMAYEAYVYKHMN-----GMNTRNIVIPTA-----SADRGSILTINHEAGY 627
Db	853 PICRDYTTNGQSATFY-GSINAQRARVECEGRDRMR-----WYKTDQIDSPPVG 899	585 YPLTYSNQGRNFEPQYOGARPEGHOPVADLPALPAAESDWTQ1RMQDGECQRY 396
Qy	691 SYIYPAMYGAYGFYQYSDDSO-----NLSDND-----716	628 NSYLF-----INCDEKTVSOGYRKSKSPVSDGQFWKERDVTREARKPEOFGV-----PYTTLVGYX 683
Db	900 QYELLSSQFGAGNTVYTPNAEIGEVQLCQTLINKPRINGSIDGAGFVRNRCEDVEGYKNNAE 959	397 CKFPOAANGDSAVEVGGDQS-----TDRCSGRDLOHWHINSANTSQOGDX 443
Qy	717 -----COLQDITKEGOLFRFLANHRANNTYMNKPHINPTE- 752	684 DPESTLSSLYPAM-----YGAYGFTYSDDSOONS-----713
Db	960 GRVWSYAIRNSEVLSRTLASQSRRELVEHRNSTHALDGRRHKSTESNKPHVNLSMEK 1019	685 :-----DNDCQLOQVDIKEG-----OLRPLRANHRANNTYMNKPHINPTE 752
Qy	753 SQPTQATLVCNNKILDTSLTTPAEGLTYYTNGOALPAKENEGCIVSYNSGKRYCLPVGQ 812	714 :-----5-OPTQATLVCNNKILDTSLTTPAEGLTYYTNGOALPAKENEGCIVSYNSGKRYCLPVGQ 812
Db	1020 GYPTOVSLSSCSD-DNGTSTT-----RTPD-ONPPDLDKLKPI-----IGQ 060	750 1 TWYRALRGDEVLRPSYQTSQCDQDVEPANGASERVNLNASRVNDS-----NKFPHYNILAMD 558
Qy	813 RSGYS 817	753 S-QPTQATLVCNNKILDTSLTTPAEGLTYYTNGOALPAKENEGCIVSYNSGKRYCLPVGQ 811
Db	1061 EYGY 1065	759 NGVPTQVRLSCSDRNGETELTREP-----QNPTADLRGPII-----IG 599
	RESULT 5	714 :-----812 QRSGY 817
Db	Q8D5P4	714 :-----812 QRSGY 817
Q8D5P4	PRELIMINARY:	714 :-----812 QRSGY 817
AC	Q8D5P4;	714 :-----812 QRSGY 817
DT	01-MAR-2003 (TREMBLrel. 23, Created)	714 :-----812 QRSGY 817
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)	714 :-----812 QRSGY 817
DE	Conserved hypothetical protein.	714 :-----812 QRSGY 817
GN	Vibrio vulnificus.	714 :-----812 QRSGY 817
OS	Bacteria; Proteobacteria; Gammaproteobacteria; Vibionales; Vibrionaceae; Vibrio; NCBI_TaxID=672;	714 :-----812 QRSGY 817
RA	Choy H.E.; Rhee J.H.; Kim S.Y.; Chung S.S.; Kim J.J.; Moon Y.H.; Jeong H.,	714 :-----812 QRSGY 817
PA	SEQUENCE FROM N.A.	714 :-----812 QRSGY 817
RC	STRAIN=CMP6;	714 :-----812 QRSGY 817
RC	STRAIN=CMP6;	714 :-----812 QRSGY 817
RA	Rhee J.H.; Kim S.Y.; Chung S.S.; Kim J.J.; Moon Y.H.; Jeong H.,	714 :-----812 QRSGY 817
RA	Choy H.E.; Rhee J.H.; Kim S.Y.; Chung S.S.; Kim J.J.; Moon Y.H.; Jeong H.,	714 :-----812 QRSGY 817
RA	Submitted (DEC-2002) to the EMBL/GenBank/PubMed databases.	714 :-----812 QRSGY 817
DR	EMBL: AE06811; AA007787; 1; -	714 :-----812 QRSGY 817
RW	Hypothetical protein; Complete proteome.	714 :-----812 QRSGY 817
SQ	SEQUENCE 875 AA; 9761 MW;	714 :-----812 QRSGY 817
Query	259 PQLNLVI---KQGNLSGRNDIKIGAPELLHTIDMGLTPRDRDFAKD-KEahr 312	Query Match 12.1%; Score 571; DB 16; Length 875;
	3 POSLBDLTNDGREGVLS---QGEIQFGGAPELIVQNDGMLIEPRNRNTMNUPLTAA 60	Best Local Similarity 27.3%; Fred. No. 8.8e-39; Mismatches 215; Indels 170; Gaps 27;
Db	313 EFOTIPVSRVNVYAPLHKEYMLPGLTIDPGNGWHSGTMRORIGKELVSGI 372	Matches 182; Conservative 91; Mismatches 215; Indels 170; Gaps 27;
Qy	61 DYFQKIPASKLVMADYTPAIFPIVTMPNGVYTDKASTGWHSGDAREAGMVTG 120	Query Match 11.1%; Score 471; DB 16; Length 875;
Db	373 DNANYGLNSTAGLGEN---SHPYVAOLAHNRSRGNAN----GIOVHGSGGGGI 421	Best Local Similarity 27.3%; Fred. No. 8.8e-39; Mismatches 215; Indels 170; Gaps 27;
Qy	121 NNAVNGLITSGAGSYQSYNRRNH-----IAHTAHTNGVYTTKTKDDLPQVIVBGSGGGGI 174	Matches 57; Conservative 14; Mismatches 35; Indels 0; Gaps 0;
Db	422 VLDSTLGNESHEKHNKGUGHYDGFGEKGSVHRSAAENNNSNTAGMDGDKKKRFIPNF---- 477	Query Match 6.2%; Score 290; DB 16; Length 957;
Qy	175 VPLEATGNEWSHEKHNKGUGHYDGFGEKGSVHRSAAENNNSNTAGMDGDKKKRFIPNF---- 477	Best Local Similarity 53.8%; Pred. No. 2.9e-10;
Db	175 VPLEATGNEWSHEKHNKGUGHYDGFGEKGSVHRSAAENNNSNTAGMDGDKKKRFIPNF---- 477	Matches 57; Conservative 14; Mismatches 35; Indels 0; Gaps 0;
Qy	478 -YPSQTNEKSLNNOQCEPP-DGHKFGDAMAGSSP--FSAAANRETMTTPNSNAILORF 532	779 LTYTYNQCOLPAKENEGCIVSYNSGKRYCLPVGORSGYSLPDATVQGEVYVDSGAKAVL 838
Db	13 LSLSRNVVQATTNEAEGCISRLNGEKYCLVGERGSYSLPSWIYAHPPDVQAPSJVSM 72	13 LSLSRNVVQATTNEAEGCISRLNGEKYCLVGERGSYSLPSWIYAHPPDVQAPSJVSM 72

Qy	73	LSDWDNLNSYRIGEVPVNNPADMKVKVAKNGQYLDLFSKPRSMRVV	864
Db	73	LSDWDNLNSIRLAVEDRYTGNEDLKVKVAKNGAYLDLFSKPRSMRVL	118
RESULT 7	Q97GW3	PRELIMINARY:	PRT: 1217 AA.
ID	Q97GW3	SEQUENCE FROM N.A.	
AC	Q97GW3;	STRAIN=ATCC 824 / DSM 792 / VRM B-1787;	
DT	01-OCT-2001 (TREMBLrel. 18, Created)	MEDLINE=21359125; Pubmed=1166286;	
DT	01-OCT-2001 (TREMBLrel. 18, Last sequence update)	Noelling J.J., Breton G., Omechenko M.V., Makarova K.S., Zeng Q.,	
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)	Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.R., Daly M.J.,	
DE	Alpha-D-glucosidase fused to unknown alpha-amylase C-terminal domain.	Tatusov R.L., Sabath F., Doucette-Stamm L., Souza P., Souza P.,	
DE	CAC225.	Bennett G.N., Kononin E.V., Smith D.R.;	
OS	Clostridium acetobutylicum.	Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum.	
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;	J. Bacteriol. 183:4823-4838 (2001).	
OX	Clostridium.	EMBL: AE007725; AAC80209.1; -.	
RN	[1]	InterPro: IPR005083; CBM 25.	
RP	SEQUENCE FROM N.A.	InterPro: IPR000322; Glyco_hydro_31.	
RC	STRAIN=ATCC 824 / DSM 792 / VRM B-1787;	InterPro: IPR000531; TonB_boxC.	
RX	MEDLINE=21359125; Pubmed=1166286;	Pfam: PF01055; Glyco_hydro_31; 1.	
RT	Noelling J.J., Breton G., Omechenko M.V., Makarova K.S., Zeng Q., Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.R., Daly M.J., Bennett G.N., Kononin E.V., Smith D.R.;	PROSITE: PS00430; TonB_DEPENDENT_REC_1; 1.	
RT	Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum.	Complete proteome.	
SQ	SEQUENCE 1217 AA;	134530 MW; 26F9D66B02688A30 CRC64;	
Query Match	79	Score 3.5%; Score 166.5%; DB 16; Length 1217;	RESULT 8
Best Local Similarity	179;	Score 3.5%; Score 166.5%; DB 16; Length 1217;	Q45780
Matches	Conservative 121;	Length 18.6%; Pred. No. 0.05%; Indels 343; Gaps 48;	PRELIMINARY;
ID	Mismatches 319;	AC 045780;	PRT: 1038 AA.
AC	Q45780;	AC 045780;	
DT	01-NOV-1996 (TREMBLrel. 01, Created)	DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)	
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)	DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)	
DE	Outer membrane protein.	DE Outer membrane protein.	
GN		GN SUSC.	
OS		OS Bacteroides thetaioraomicron.	
OC		OC Bacteroides; Bacteroidetes; Bacteroides (class); Bacteroidales; Bacteroidaceae; Bacteroides.	
OC		NCBI_TaxID:818;	
RN		RN [1].	
RP		SEQUENCE FROM N.A.	
RC		STRAIN=BT5482;	
RX		MEDLINE=961146534; PubMed=8550519;	
RA		Reeves A.R., D'Ella J.N., Frias J., Sayers A.A.;	
RT		"A Bacteroides thetaioraomicron outer membrane protein that is essential for utilization of malicooligosaccharides and search.";	
RL		J. Bacteriol. 178:823-830 (1996).	
CC		CC SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).	
DR		DR InterPro: IPR01395; Aldo_ket_red.	
DR		DR InterPro: IPR00531; TonB_boxC.	
DR		DR Pfam: PF00593; TonB_dep_Rec_1.	
DR		DR PROSITE: PS00063; ALDOKETO_REDUCTASE_3_1.	
KW		KW Membrane; Outer membrane; Receptor; TonB box.	
SQ		SQ SEQUENCE 1038 AA; 115886 MW; DD45CE382DC5E1 CRC64;	
Query Match	231	Score 3.4%; Score 151; DB 2; Length 1038;	409
Best Local Similarity	276	Score 3.4%; Score 151; DB 2; Length 1038;	445
Matches	SGELEN-NRITYAQHWSAELPAHIVPGLNLKOGNLSGRLND-----	SGELEN-NRITYAQHWSAELPAHIVPGLNLKOGNLSGRLND-----	
ID	276	SGELEN-NRITYAQHWSAELPAHIVPGLNLKOGNLSGRLND-----	
AC	276	SGELEN-NRITYAQHWSAELPAHIVPGLNLKOGNLSGRLND-----	
DT	01-NOV-1996 (TREMBLrel. 01, Created)	SGELEN-NRITYAQHWSAELPAHIVPGLNLKOGNLSGRLND-----	
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)	SGELEN-NRITYAQHWSAELPAHIVPGLNLKOGNLSGRLND-----	
DE	Outer membrane protein.	SGELEN-NRITYAQHWSAELPAHIVPGLNLKOGNLSGRLND-----	
GN		GN SUSC.	
OS		OS Bacteroides thetaioraomicron.	
OC		OC Bacteroides; Bacteroidetes; Bacteroides (class); Bacteroidales; Bacteroidaceae; Bacteroides.	
OC		NCBI_TaxID:818;	
RN		RN [1].	
RP		SEQUENCE FROM N.A.	
RC		STRAIN=BT5482;	
RX		MEDLINE=961146534; PubMed=8550519;	
RA		Reeves A.R., D'Ella J.N., Frias J., Sayers A.A.;	
RT		"A Bacteroides thetaioraomicron outer membrane protein that is essential for utilization of malicooligosaccharides and search.";	
RL		J. Bacteriol. 178:823-830 (1996).	
CC		CC SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).	
DR		DR InterPro: IPR01395; Aldo_ket_red.	
DR		DR InterPro: IPR00531; TonB_boxC.	
DR		DR Pfam: PF00593; TonB_dep_Rec_1.	
DR		DR PROSITE: PS00063; ALDOKETO_REDUCTASE_3_1.	
KW		KW Membrane; Outer membrane; Receptor; TonB box.	
SQ		SQ SEQUENCE 1038 AA; 115886 MW; DD45CE382DC5E1 CRC64;	
Query Match	302	Score 3.4%; Score 151; DB 2; Length 1038;	445
Best Local Similarity	302	Score 3.4%; Score 151; DB 2; Length 1038;	445
Matches	VLNTRASKQIFIDNCFCLDFPEKKKGODNTGEPFKWNTDNFDQNLCKMD-SKGKLM	VLNTRASKQIFIDNCFCLDFPEKKKGODNTGEPFKWNTDNFDQNLCKMD-SKGKLM	
ID	302	VLNTRASKQIFIDNCFCLDFPEKKKGODNTGEPFKWNTDNFDQNLCKMD-SKGKLM	
AC	302	VLNTRASKQIFIDNCFCLDFPEKKKGODNTGEPFKWNTDNFDQNLCKMD-SKGKLM	
DT	01-NOV-1996 (TREMBLrel. 01, Created)	VLNTRASKQIFIDNCFCLDFPEKKKGODNTGEPFKWNTDNFDQNLCKMD-SKGKLM	
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)	VLNTRASKQIFIDNCFCLDFPEKKKGODNTGEPFKWNTDNFDQNLCKMD-SKGKLM	
DE	Outer membrane protein.	VLNTRASKQIFIDNCFCLDFPEKKKGODNTGEPFKWNTDNFDQNLCKMD-SKGKLM	
GN		GN SUSC.	
OS		OS Bacteroides thetaioraomicron.	
OC		OC Bacteroides; Bacteroidetes; Bacteroides (class); Bacteroidales; Bacteroidaceae; Bacteroides.	
OC		NCBI_TaxID:818;	
RN		RN [1].	
RP		SEQUENCE FROM N.A.	
RC		STRAIN=BT5482;	
RX		MEDLINE=961146534; PubMed=8550519;	
RA		Reeves A.R., D'Ella J.N., Frias J., Sayers A.A.;	
RT		"A Bacteroides thetaioraomicron outer membrane protein that is essential for utilization of malicooligosaccharides and search.";	
RL		J. Bacteriol. 178:823-830 (1996).	
CC		CC SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).	
DR		DR InterPro: IPR01395; Aldo_ket_red.	
DR		DR InterPro: IPR00531; TonB_boxC.	
DR		DR Pfam: PF00593; TonB_dep_Rec_1.	
DR		DR PROSITE: PS00063; ALDOKETO_REDUCTASE_3_1.	
KW		KW Membrane; Outer membrane; Receptor; TonB box.	
SQ		SQ SEQUENCE 1038 AA; 115886 MW; DD45CE382DC5E1 CRC64;	
Query Match	357	Score 3.4%; Score 151; DB 2; Length 1038;	445
Best Local Similarity	357	Score 3.4%; Score 151; DB 2; Length 1038;	445
Matches	GTMQRRI----GKELVSH-	GTMQRRI----GKELVSH-	
ID	357	GTMQRRI----GKELVSH-	
AC	357	GTMQRRI----GKELVSH-	
DT	01-NOV-1996 (TREMBLrel. 01, Created)	GTMQRRI----GKELVSH-	
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)	GTMQRRI----GKELVSH-	
DE	Outer membrane protein.	GTMQRRI----GKELVSH-	
GN		GN SUSC.	
OS		OS Bacteroides thetaioraomicron.	
OC		OC Bacteroides; Bacteroidetes; Bacteroides (class); Bacteroidales; Bacteroidaceae; Bacteroides.	
OC		NCBI_TaxID:818;	
RN		RN [1].	
RP		SEQUENCE FROM N.A.	
RC		STRAIN=BT5482;	
RX		MEDLINE=961146534; PubMed=8550519;	
RA		Reeves A.R., D'Ella J.N., Frias J., Sayers A.A.;	
RT		"A Bacteroides thetaioraomicron outer membrane protein that is essential for utilization of malicooligosaccharides and search.";	
RL		J. Bacteriol. 178:823-830 (1996).	
CC		CC SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).	
DR		DR InterPro: IPR01395; Aldo_ket_red.	
DR		DR InterPro: IPR00531; TonB_boxC.	
DR		DR Pfam: PF00593; TonB_dep_Rec_1.	
DR		DR PROSITE: PS00063; ALDOKETO_REDUCTASE_3_1.	
KW		KW Membrane; Outer membrane; Receptor; TonB box.	
SQ		SQ SEQUENCE 1038 AA; 115886 MW; DD45CE382DC5E1 CRC64;	

QY	39	NDLOGSLAAEVKFAQSILPAHPKGEDSQPHITSLRKSTLLVRR---PVKADDKTP---v	91	AC Q46034; DT 01-NOV-1996 (TREMBrel. 01, Created) DT 01-NOV-1996 (TREMBrel. 01, Last sequence update) DT 01-MAR-2003 (TREMBrel. 23, Last annotation update)
Db	131	NDMTGGSVA-----IKPDELSKG-----ITNAQDMSLGSKTAGVSYTISNDTPGGGA	177	Toxin_B
QY	92	QEARDNNKKILGLTLTYPSSPLPOTIYHLDGVPEGGIQDFTPHNGTKKLINTAEVNK-	149	Clostridium difficile.
Db	178	QIRIGGSSL-----NASNDPLIVDGL---AID---NEGIKGANGL-SMVNPAD	221	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiales; Clostridaceae;
Db	150	-----LSDAGGSSIHSHLTNNALVEHTANGRWRDILPQGPDLLEGKQVRFYSSAGYS	203	Clostridium, NCBI_TAXID=1496;
Qy	222	IETLTVLKDDASATAYGRASNVIIITTKKGK-----NGQAPSVTYNGSVSFS	270	RN [1] SEQENCE FROM N_A.
Qy	204	STVFYGDRTKVLTUSVGNTLFLFVQWFRSGELENRNRITYAOHWSAELPA-----HW	236	RA STRAIN-isolate_1470;
Db	271	KT-----QKRDVLVSGDEYRA-----YANOLWGDKIPADLGATANDW	307	RA *Closing in on the toxic domain through analysis of a variant
Qy	257	-----IVPQLNLVTRQG--NLSGRRLNDIKIGAPGEPLLHTIDIGMLTTPRDRDFDAK	306	RT Clostridium difficile cytoxin B.*;
Db	308	QDQIFRTAVSTDHHVSINGGFKNLQPYRS-----LCYTDNGIVKTSNEF-RFTA	356	RL Mol. Microbiol. 17:313-321(1995).
Qy	307	DREAHNEYFO---TIPYRSRMV---NNYAPLHLKE---VMLPTGEELLTDMDP---GNNG	353	DR Eichel-Streiber C., Meyer zu Heringdorf D., Habermann E., InterPro; IPR002479; CW_binding.
Db	357	SVNLAPSFFEDHLKPNINAKFMGNKRNADRSRCRYWRAIAIDPRTPVSNEDYQFTGCV	416	DR InterPro; IPR001950; TIF_SUI1; Pfam; PF01473; CN_binding_1; 18.
Qy	354	WHSGTMRQRIGKELVSHGIDNANYGLNSTAGLE-----NSHPYTTVAQLAHNNSRC	404	DR PROSITE; PS1118; SUT1; 1. DR POSITIVE; PS1118; SUT1; 1. SQ SEQUENCE 2367 AA; 269186 MW; EP9823DAE70427F3 CRC64;
Qy	417	WQN-----INSTIGFSNDWKYTISNPNSPQNPLAALELKNDKG	454	Qy Score 159.5 ; Pred. No. 0.45 ; DB 2 ; Length 2367;
Db	492	TIVSPYSGNNYGYWNGDVTOKYKYNLSNYTYYQYI-KSLGAND-----FDIMV	538	Best Local Similarity 18.4% ; Pred. No. 0.45 ; DB 2 ; Length 2367;
Qy	508	GGSPFSAANRFTMYTPNSALLQREFENKAVALDPSRSSTGFSK-----WNADTOEMEPYEHTI	564	Matches 185; Conservative 149; Mismatches 321; Indels 353; Gaps 49;
Db	455	N-----SNDFGVGNVDYDVKFHFLPDLRHLASTGGEYAESTQT	491	Qy 27 NSAYLFNTSNTSOPINDLOGSLAAEVKAQSOQILPAHPKEGDSQPHLTSLRKSLULLYRPVKAD 86
Qy	452	SYHRSANNNNSTWGRGD-----KKRTEPNNPQSTQNEKCLNQEQPFDGHKFGFDA	507	DB 1454 NSLOQNKNPYFSYDSEGK-----BNGFPINGSTREG-----LFVSPLPDVLSKVMID 1501
Db	499	-----QHFRN-----GPEEGQWDSTOE-----PHDA	566	Qy 87 DKTP-----VYOEARDNNNKKL-----TILTLYPPSSLPDTIYHLDGV 124
Db	562	565 DRAEQITASYNLESESKMA-----ELMAEAYAVKVKHMNGNWTRNYYIPTASADNRGSILT	620	DB 1502 DSXPSFGYSSNNEURKDVKVITKDNISLTLQDEKTIKLNSYHLD-----1559
Qy	567	KREOQAYATATRNLYSFGRILNYSLLNRYLFETMRWDGS-----SREFSKDNRW---	615	Qy 125 PEGIDDEFPHNGTKKIINTVAYENKVLSDASGSSSIHSHTNRLALVEHTAANGRWRDIVLP
Db	621	INHEAISYNSLLEINGDEKVTQSQGYKSFTSYQDFERDVTDEARKPQFGPVTTLV	680	DB 1560 -ESG-----VAETLKFMRNKG-----TNTSDSMSFILESMN---IKSIVF- 1596
Qy	616	-----GTFPSLALGWK-----IKEENFLKDVNLSLSDKLKR-----LGNGITGQQ	654	Qy 185 QGPDPLEGKMRVFRYSSAGY-----SSTVFKG-----DRKVTLISVGNT 220
Db	681	GYDPEBTLSYYIYPAMYGAVF---TYSDDS---QNLSDNDQCQLQVDTKEQRFLRANHR	736	DB 1597 ---NFLQSNKIFLDANFISGTTSIGQFEETCDENNINIQPYFIKNTLETNLYVGNER 1653
Qy	655	NGDDFAKLPLVYVNEYAQYPFGTYYSQVQDPLRQVWV	714	Qy 221 L-LFKYNGQWFSGELENRTRIYAO-HWSAELPAWIYPLNLYIKGONLUSG-----272
Db	737	ANTVYANKKEFHINVPTEQOPTAOATLYCNKILDTKSLTPAPEGLTY-----TVNCQ	786	DB 1654 QNMIVEPVNPDYLDLSDGSDISSTVINFQSYLYGIDSCVNUKVVISPNITDEINTPYETNN
Qy	715	771 TIGGIDGYF-----RKTMCVTAIRSPMNLQCPDDEYRFTGKLXRGFSINAK	763	Qy 273 -----RLNDIKI-----GAPGELLHTIDIGMLTPDRDF-DFAK 306
Qy	787	ALPAKE-----NEGIVSYNSG--KRYCPLVG-----	811	DB 1714 TYPEIVILDANYNEKINNNINDLSIRYVWSNDGDEFLMSTSEEVKVSVQYKRFVNFK
Db	764	772 ANTMTKKEFHINVPTEQOPTAOATLYCNKILDTKSLTPAPEGLTY-----TVNCQ	786	Qy 339 -----PTIGELD-----MDPGMGHHSQGMNRQIGKELYSHG 372
Qy	812	-----QRSQYSLPDWIVQEVTYDSGAKAVLSDWNLSYNRIGEFVQGNY	858	Db 1831 VSGLIYINDSLYFKPPVNNLITGFYVGDKEYFNPINGGAAS-----IGETI-----1 1882
Qy	824	FYYSRGNNKTYQAHKGYAANSFPTYQQVDFDENGKPI-----ENMFDRNGN-GTID	874	Qy 373 DNANYGLNSTAGLGENSHPPVYQAALARNSRGNY-----ANGIOVHGSSGGGTYLDST
Db	859	PADMKKVKAWS-----QYLDFSKPRSMR	882	DB 1883 DDNNYFPROG-----VQZQGTFSTEDGLYDGLVLSYNEKFYINNFGMM
Qy	875	SGDDKYIYKPKADGVLMCLTSKMQYKNFDFSFSLR	908	Qy 428 LGNE---PSHEYGHNYGLGHYDGFKGSVRSAAENNSTCW---DGDKRKFPIPNFYSQ
RESULT 9	1	-----TNAKSCLNQCOEPFDCHKTFGDMAGGSPSAAANRF--TATTPNSAIIQRF	EN 535	DB 1919 LEGBAIDTKLKLIDILEYFEDNYRGAVE-----WKELEMHYSFSP-----1962
ID 46034	2	PRELIMINARY;	3	Db 1963 -----TGKAFAKGLNQIGDDKYYFNSDGVWOKGFVNSINDN 1996



RC	STRAIN=ATCC 824 / DSM 792 / VRM B-1787;	Db	681 NYGEKKITDYYVDAFPGNEASSFDYDDDGTSNYENGSSYFDQKMTLERAQLKSVQFN 740
RX	MEDLINE=21359325; PubMed=11466286;	Qy	674 VPTTLYVYDPEGTSSYTP-----NMGAFTYDDSONLSDNDC 717
RA	Noelling K.M.V., Omeilchenko M.V., Makarova K.S., Zeng Q., Breton G., Gibson R., Lee H.M., Dubois J., Olu D., Hitti J., Wolf Y.I., Tatusov R.L., Sabatte F., Poucette-Stamm L., Soucaille P., Daly M.J., Bennett G.N., Kooin E.V., Smith D.R.;	Db	741 ISPKT-GYY-KSDLKNTYKMHVKSSGDVTVGRRRTRYASY----DELNNAQEGY 791
RA	"Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum."	Qy	718 QLOQDTKEQLRFLR-ANHRANNTVANKFHINPTEQPTQATLVCNKKIL---DTKSL 772
RT	J. Bacteriol. 183:4323-4328 (2001).	Db	792 WQTDIYGSVYIYKVSAGHORN----INV-----CQVOLTAKADVK- 830
RL	EMBL: AE07786; AAC0833..1..;	Qy	675 TPAGEGLTYTNGOALPAKEN-----EGCIVSYNSGKRYCLPGSORSGYSLPDWI 822
DR	InterPro: IPR005085; CBM_25;	Db	831 ----GTYT-SPOVKSLASDPNAIYTUDGTAPTVNPK-----YTGP-- 870
DR	InterPro: IPR000322; Glyco_hydro_31.	Qy	823 VGOENYVDSGAKAVLISDADNLSNRIGEVGYNPADMKVKWANGGYLDSESKPRSM 882
DR	InterPro: IPR00531; TonB_boxC.	Db	871 ----ITDSKTLKFIVR-----ANGESDV----FTEQYTY----IK 903
DR	Pfam: PF01223; CBM_25;	Qy	883 VVYK 886
DR	Pfam: PF01055; Glyco_hydro_31..1..;	Db	904 VVYK 907
KW	PROSITE: PS0430; TonB_DEPENDNT_REC_1..1..;		
SEQUENCE	Hydrolase; Complete proteome; Sequence: 1157 AA; 129411 MW;		
SQ	1DB1728DB33F7578 CRC64;		
	Best Local Similarity 19.3%; Score 156.5%; DB 16; Length 1157;		
	Matches 198; Conservative 120; Mismatches 335; Indels 371; Gaps 56;		
Qy	83 YKADDKTPVOYEAARDDNNK----ILGTLTLYPPSSLPLDTIYLHDQVPEGIDFTPHNGTK 138	RESULT 12	
Db	35 VADTNKVNNKNSQKKPHAKLNGNTLKKKGKDETIRICEPQVFKVDYKP-NGKS 93	QPF931	PRELIMINARY:
Qy	139 KLTINTVAEVNLKSDASGSSSIHLTNALVIEHTANGRKYRDI-----YLPOQPDYL 189	ID QPF931	PRT; 2367 AA.
Db	94 SKDTLVDPNK-KWSTONNIVSSDKIDPDMV --ITTKKAVLKIINKEDLSTLIVY ---DL 144	AC QF931..	
Qy	190 EGSKMRYREVSSAGYSSYTF--YGDRTKVTLSVNTLFLKFYNGQWFR---SGELEN-NRI 241	RC 01-MAR-2001 (TREMBLrel. 16, Created)	
Db	145 QKLRLKQDSTASKTAAFTHSGDR -----FYGNGYNQFEDSKGMNRGTS 193	DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)	
Qy	242 TYQHIMSAELPAHWIYPGLNLVIKQGNLSSLRND-----IKIGAPGL 285	DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)	
Db	194 VYAGYQGHCGSPFWNSNDGYGLVSDGGSTTIGOTSLKYDGISKTDTDYVMYGNPKEL 253	DE Cytotoxin B.	
Qy	286 LHTIDIG---MLTIPDRF -----DFKDKEARHEY_FQTIPVSRI 324	TCDB Clostridium difficile.	
Db	254 LSPESDVSGKAPEKPKWANGTNTQGWDNSLISGTGNDEARLKSNTYTRSKOLPIDNRC 313	OC Bacteria: Clostridia; Clostridiales; Clostridaceae;	
Qy	325 V-----NNYAPHLKEYMLP--TGELLTDMDPGNGWHSGTMQRQI-----GRE 366	OC Clostridium	
Db	314 LDPPDKWKGQDNTGEGFWNTDRFDSONGQKXAYMD-SKGKMTGTMKPKLADSKQGRY 372	OC NCBI_TAXID=1496;	
Qy	367 LVSHG-----IDNANYGLNSTAGLGENSEPPTVAAQLAHNNSRGNYANC 409	OX RN	
Db	373 VTSKGMWLPGDSASDVGSKGMENYFAL-----PQV-----RKWMWN 412	RP SEQUENCE FROM N_A.	
Qy	410 IOYHGGSGGGIVTLDSTLGNFESHEHGNVGLHY-----VDFGFKSVRSAAENN 460	SPRAINS5340;	
Db	413 IQ--WNEDECENYTFGNFGNMANMERA1YDG-----ORHEN 456	RA RXA	
Qy	461 NSTWG-----WDGDKKRFRIDNPFTPSOTNEK--SCLNNOQCQFPFDGHK 500	RA Sambol S.P., Metzger M.M., Leyerly D., Gerding D.N., Johnson S.;	
Db	457 QRTWSLNRNYYAGAQRSYGMGSD---ISTGFDMDAQNPERMSAVN-----LGEAK 506	RT "Toxin gene analysis of a variant strain of clostridium difficile that causes human clinical disease."	
Qy	501 FGFD--AHAQSP---PSAABRFTMTPNSNALLORFF--ENKA---VFDSSRSTGFS 548	RT Infect. Immun. 68:5480-5487 (2000).	
Db	507 WGMDTGGFNGDPTPENYARREFSATP---IPRVHGDKNKRYPWAEGSTREATAK 561	DR ENBML: AF217292; AGI18011..1..;	
Qy	549 KNNADTOEKEPYHTIDRAEQTAA-----SYNLESESKMELMAEYAYKHM 596	DR InterPro; IPR002479; CW_binding.	
Db	562 KAMOLRYTLLPYIYSYDRAOSQGLGLYRSILMEYPMDSNLANDKEDAWMFGDYMVLSPVV 621	DR IPRO011950; TIF_SDR1.	
Qy	597 WNGWTRNLYIP-----TASADNRGSLTINHEAGNSY----LFING-----	DR PROSITE; PS00118; SUTI1..1..;	
Db	622 QEGTSKSYLPGEGNWIDYTGREYTG-OTINYAVDSTNSDIPFKSGSAILPTQDFE 680	DR POSITIVE; PS00118; SUTI1..1..;	
Qy	636 ---DEKVYSSQYKSF-----VSDGQFWKERDVDTREARKPEQFG 673	DR 125 PEGSGLDFPHNGSTKKLINTVAFVNKLSDAGSSISIHSLTNNALVEHTANGRMVYDYL 184	
Db	-----	DR 1550 -ESG-----VAETLKFMRKGS---TNTSDSMSPLESMN---DRKVTLISVGNT 220	
Qy	-----	DR 185 QGPDPLEGKMYRFVSSAGY--SSTVYFG-	
Db	-----	DR 1597 ---NFGSNLKEFLDAMTIISSCTTSISQEEFICDENNIQPYFKEINTLTYVGR 1653	
Qy	-----	DR 221 L-LFKYVNRQWPSGELENNRITYAQ-HWSAELPAHWIVPGLNLVIKQGNLSG-----272	

Db	1654	QNNIVEPNYDDDDSGDISSTVINFQSKYLYGIDSCVNYKVVSPNIYTDEINTPVYETNN	1713	Qy	273	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	273	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	273	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Db	1714	TYPEVIDANYNEKINVNINDSLTRVNSNDGDFILMSTSEENKVSOVKIRFVNFK	1773	Qy	1714	TYPEVIDANYNEKINVNINDSLTRVNSNDGDFILMSTSEENKVSOVKIRFVNFK	1773	Db	1714	TYPEVIDANYNEKINVNINDSLTRVNSNDGDFILMSTSEENKVSOVKIRFVNFK	1773	Qy	307	DKE-BHREYF-----QTIPYSRMIVNNYAPHLKEYML-----	338
Db	1774	DKTFLANKLSFNFSDKQDVPVSEIIL-SFTPSYYEDLGIDGLVSLYNEKFTINNNFGNM	1832	Qy	339	-----PTCELTID-----MDPGNGWHSGTMQRIGKELVSHGI	372	Qy	339	-----PTCELTID-----MDPGNGWHSGTMQRIGKELVSHGI	372	Qy	1833	VSGLIYINDSILYYFKPVNNITL-SFTPSYYEDLGIDGLVSLYNEKFTINNNFGNM	1882
Db	1833	VSGLIYINDSILYYFKPVNNITL-SFTPSYYEDLGIDGLVSLYNEKFTINNNFGNM	1882	Qy	1833	VSGLIYINDSILYYFKPVNNITL-SFTPSYYEDLGIDGLVSLYNEKFTINNNFGNM	1882	Qy	373	DNNYGINSTAGLGENSHPYVIAQLAHHNSRGNY-----ANGIQHGGSSCGGYTLDST	427	Qy	373	DNNYGINSTAGLGENSHPYVIAQLAHHNSRGNY-----ANGIQHGGSSCGGYTLDST	427
Db	1883	DDKNYYFNQS-----VLTQTYFVSTDGFKFYFAPAN-----TLDEN	1918	Qy	1883	DDKNYYFNQS-----VLTQTYFVSTDGFKFYFAPAN-----TLDEN	1918	Qy	1883	DDKNYYFNQS-----VLTQTYFVSTDGFKFYFAPAN-----TLDEN	1918	Qy	428	LGN-----FSHEVGHNTGLGHYDGFGRGSVHRSAAENNINSTGW-----DGDKRKFIPNFYPSO	481
Db	1919	LEEAIDFTGKLIDENITYFDNYNSRAVE-----WKEDELGEHYFSE-----	1962	Qy	1919	LEEAIDFTGKLIDENITYFDNYNSRAVE-----WKEDELGEHYFSE-----	1962	Db	1919	LEEAIDFTGKLIDENITYFDNYNSRAVE-----WKEDELGEHYFSE-----	1962	Qy	482	TNEKSCLNQCOEPFDGHKFDFAMACGSPFAAANRF-----TMTPNNSAIIQRF	535
Db	1963	-----TGKATFKGLNQIGDDKYYFHSIDGVAQKGFSINDN	1996	Qy	1963	-----TGKATFKGLNQIGDDKYYFHSIDGVAQKGFSINDN	1996	Db	1963	-----TGKATFKGLNQIGDDKYYFHSIDGVAQKGFSINDN	1996	Qy	536	KAYEDSR-----STGFSK-----WNADTOQEMPEHTIDRAEQTIASVNELESKMAELMA	587
Db	1997	KHFDDSGVMKVYSTEIDGKHYFAHNEDGNERGEESIS	2056	Qy	1997	KHFDDSGVMKVYSTEIDGKHYFAHNEDGNERGEESIS	2056	Db	1997	KHFDDSGVMKVYSTEIDGKHYFAHNEDGNERGEESIS	2056	Qy	588	EYAVVKVHMNGWTRNIIYIPTASADNRGSLTINHEAGYNSYLTINGDEKVVSGYKK	647
Db	2056	-----NPNKIIY---FDDFTAVYWKDQLEDSKYYF-----DEUTAEAYIGLS	2103	Qy	2056	-----NPNKIIY---FDDFTAVYWKDQLEDSKYYF-----DEUTAEAYIGLS	2103	Db	2056	-----NPNKIIY---FDDFTAVYWKDQLEDSKYYF-----DEUTAEAYIGLS	2103	Qy	2057	YSGFL-----NPNKIIY---FDDFTAVYWKDQLEDSKYYF-----DEUTAEAYIGLS	2103
Db	2104	LINDQQYFNDG1-----MQYGVFTINDKFVIFSDS2IES-----GVQNI	2145	Qy	2104	LINDQQYFNDG1-----MQYGVFTINDKFVIFSDS2IES-----GVQNI	2145	Db	2104	LINDQQYFNDG1-----MQYGVFTINDKFVIFSDS2IES-----GVQNI	2145	Qy	2146	DDNPFYIDDNGIVIQGVFTDSQCKYFAPA-----NTVNDNITY-----	2183
Db	2146	DDNPFYIDDNGIVIQGVFTDSQCKYFAPA-----NTVNDNITY-----	2183	Qy	2146	DDNPFYIDDNGIVIQGVFTDSQCKYFAPA-----NTVNDNITY-----	2183	Db	2146	DDNPFYIDDNGIVIQGVFTDSQCKYFAPA-----NTVNDNITY-----	2183	Qy	2146	DDNPFYIDDNGIVIQGVFTDSQCKYFAPA-----NTVNDNITY-----	2183
Db	2184	QGAEVYSGLVRVEDDVYFGETYTIENGWIKDMENE-----SDKYYFDETKACKGI	2236	Qy	2184	QGAEVYSGLVRVEDDVYFGETYTIENGWIKDMENE-----SDKYYFDETKACKGI	2236	Db	2184	QGAEVYSGLVRVEDDVYFGETYTIENGWIKDMENE-----SDKYYFDETKACKGI	2236	Qy	2184	QGAEVYSGLVRVEDDVYFGETYTIENGWIKDMENE-----SDKYYFDETKACKGI	2236
Db	2237	NLDDI-----KYYFDEKGIMRIGLISFENNNYYFNENGMQQGYIINIED	2281	Qy	2237	NLDDI-----KYYFDEKGIMRIGLISFENNNYYFNENGMQQGYIINIED	2281	Db	2237	NLDDI-----KYYFDEKGIMRIGLISFENNNYYFNENGMQQGYIINIED	2281	Qy	2237	NLDDI-----KYYFDEKGIMRIGLISFENNNYYFNENGMQQGYIINIED	2281
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306
Qy	2237	-----RNDIKIT-----GAEFLLHTIDGMUJTPRDF-DAK	306	Qy	2237	-----RNDIKIT									



Qy 104 ..... GTI,TLYPPSSSLPD ..... -TIVHLDGV 124  
 Db 444 VTIITIDNGNPIGQTSDAQNNSSTPGSOLPDTWVNNVARDAAQNSSPATTSITVDCY 503  
 Qy 125 PEGGIDETPHNGTKKINTVAYEVK ..-LSDASGSITHSHLTNNNALEVTHTANGRWRDI 181  
 Db 504 AENAPVVEPSNSE- LSGTAPGSSVTLTGNGNPIGOTTAD .....-ANGNW-- S 55C  
 Qy 182 YLPOGPDLGEKMRVEYS- -SAGYSS .....-TVFYGDRKVTLVSGNTLLEKYNGQF 231  
 Db 551 FFPSTPIPDPGTWVNNVARDAAQNSSPASVYDAAPATTVDPNTGTT----- 600  
 Qy 232 RSGELENNRITYAQHINSAELPAHWIVPGNLVVIQOGNI.SGRNLNDIKIGAPCELLHTID 291  
 Db 601 -SGTAE .....-PGSSEVTLDPGN----- 619  
 Qy 292 IGMLTTPDRDFEAKDEAHBEYFQTFIPVSSMIVNNVAPLILKEVALPTGELLTD----- 346  
 Db 620 IGOVTA-----DGSQNTWFTPSTPLPRTVYNATAID-----PSQNAASSPASVY 664  
 Qy 347 .....-MDPGNGGHSGSTMQRIGKELYSHGIDDNANTYGLNSTAGLGENSHPYVVAQL 397  
 Db 665 DAVAPATPVNFSNGTTLSGT -----AEPAVATVLTDG NGNP- IGQV 705  
 Qy 398 AAHNSRGNYA .....-NGIQVH .....-GGSGGGGIVTLD----- "TL 428  
 Db 706 TADGS- -CNWFSTPTPLPNTVNNATDAGSNTSAGSSVTVDSYAPATPVINPSNGTTL 764  
 Qy 429 GNEFSHEVGHNTGLGHYDGFKGSVRSAAENNNSTMWDGDJKKRETPNFTPS -QTNEKSC 487  
 Db 765 SG -TAEPGSSYTL-- TDGNCNPICQVTAQCSGNM-----SFTPSPLADGTV 808  
 Qy 488 LNNOCQEPFDGHKFGRFDAMAGSSPFSAANRRTMYPNNSAIIORFFENKAVFDSSSSTGF 547  
 Db 809 VNATATDP-----AGNTSGCGSTTVTDGVAPTTPV-----NLSNGSSLGTA - 850  
 Qy 548 SKNNADTOEMEPYEHNTDRAEQITASVNELESSESKHAELMAEYAVVYHMNINGNNWTRNIYI 607  
 Db 851 -----EPOS----- -TVILTGDGNPIREVAD----- -GSQNW-----YT 880  
 Qy 608 PTASADNRGSILTINHEAGYNSYLINGDEKVSOQYKKSTVSDQFWKERDVDTREAR 667  
 Db 881 PСПIANSTVVNVQAQDAAGNS-----SPEASV----- -VDSQAPA 916  
 Qy 668 KPOFGYPTVLTGYDPEGTSSYIYPMNGAYGFTYSDDSQNLSDNDQLOVDTKEQ 727  
 Db 917 APVYNPSNGTUSLGTAAPEGATV-----TLTDNGNPDG-----QVTDAD-GSGN 958  
 Qy 728 LRFLPLANHANNNTVANKPHINVPTEOPTQATLVCNNKILDTSLTPAPELTYVNGQA 787  
 Db 959 WSSTPGTPLANGTVVN-----ATASDPTGNTSAPASTV-----STAPAAP-VVNPNSNGAE 1009  
 Qy 788 LPARENECIVSYNSGKRYCLPVGQRSGYSPDW 821  
 Db 1010 ISGTAEGATVLTGSGN- -PIGOVTAADGSGNW 1041

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 8, 2003, 16:20:57 ; Search time 21.2134 Seconds  
(without alignments)  
(3912.317 Million cell updates/sec)

Title: US-10-002-309B-2\_COPY\_24\_886  
Perfect score: 4597

Sequence: 1 ADNNSATYFNTSQPINDLQG.....AWNGQYLDLDFSKPRSMRVYTK 863

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters:

283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR-76;\*

1: Pir1;\*

2: Pir2;\*

3: Pir3;\*

4: Pir4;\*

RESULT 1

T42131

probable toxR-regulated lipoprotein tagA - Escherichia coli Plasmid pO157

C;Species: Escherichia coli

C;Date: 03-Dec-1999 \*Sequence\_revision 03-Dec-1999 #text\_change 17-Nov-2000

C;Accession: T42131; T00210

R:Burland, V.; Shao, Y.; Perna, N.T.; Plunkett, G.; Sofia, H.J.; Blattner, F.R.

Nucleic Acids Res. 26:4106-4204, 1998

A;Title: The complete DNA sequence and analysis of the large virulence plasmid of E. coli K12

A;Reference number: 222068; PMID:98391744; PMID:9722640

A;Accession: T42131

A;Molecule type: DNA

A;Residues: 1-898 <BUP>

A;Cross-references: EMBL:AF074613; PIDN: AAC70099.1

A;Experimental source: strain EBL933; serotype O157:H7

R:Makino, K.; Ishii, M.; Yasunaga, T.; Hattori, M.; Yokoyama, K.; Yatsudo, H.C.; Kubo S.; Shinagawa, H.

DNA Res. 5, 1-9, 1998

A;Title: Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic

A;Reference number: 214127; PMID:98290540; PMID:9628576

A;Status: preliminary; translated from GB/EMBL/DDJB

A;Molecule type: DNA

A;Residues: 569-898 <WAK>

A;Cross-references: EMBL:AB011549; PIDN:BA31757.1; PMID:di1032718

A;Experimental source: strain EHEC O157:H7; substrate RIMD 0509952

C;Genetics:

G;Genome: Plasmid pO157

A;Note: L7031

Query Match 100.0%; Score 4597; DB 2; Length 898;

Best Local Similarity 100.0%; Pred. No. 5.7e-292;

Mismatches 0; Conservative Matches 863; Indels 0; Gaps 0;

QY 1 ADNNSAIYFNTSQPINDLQGSLAAEYKFAQSQTLPAPKEGSQPHLSURKSLLIVRPy 60

DB 36 ADNNSAIYFNTSQPINDLQGSLAAEYKFAQSQTLPAPKEGSQPHLSURKSLLIVRPy 95

QY 61 KADDKTPVQEARDDDNNNKLGTLTLYPPSSLPTDTYHLDGYPEGGIDFTPHNGTKIINT 120

DB 96 KADDKTPVQEARDDDNNNKLGTLTLYPPSSLPTDTYHLDGYPEGGIDFTPHNGTKIINT 155

QY 121 VAENKLSDASGSSISHLSTLNALVEIHTANGRWDYLQPGDLEGKAYRFSSAGYS 180

DB 156 VAENKLSDASGSSISHLSTLNALVEIHTANGRWDYLQPGDLEGKAYRFSSAGYS 215

QY 181 STVFYGDKEYTSLVSGNTLFLKYINGQWRSGELENNRTYAHINSAELPAHWIVPGLNL 240

DB 216 STVFYGDKEYTSLVSGNTLFLKYINGQWRSGELENNRTYAHINSAELPAHWIVPGLNL 275

QY 241 VIKQGNLSGRNDIKIGKAGELLRTIDIGMUTTPDRDFAKDEKAHREYFQTIPVSRM 300

### ALIGNMENTS

30	133.5	2.9	2660	2	E85822
31	133	2.9	4199	2	S76412
32	132	2.9	815	2	H82345
33	132	2.9	2273	2	T09083
34	130.5	2.8	773	2	JB0387
35	130	2.8	813	2	A40601
36	129.5	2.8	2186	2	T13169
37	129	2.8	1250	2	D91018
38	128.5	2.8	1004	2	B25039
39	128.5	2.8	1051	2	G34316
40	128.5	2.8	1238	2	A64596
41	128.5	2.8	1939	2	D97316
42	128	2.8	1250	2	F85862
43	128	2.8	1487	2	AG2560
44	127.5	2.8	936	2	S57637
45	127.5	2.8	1067	2	D75625

Db	276	VITQGNLSGRLNDIKIGAPGELLHTIDGMITTPRDRFDEAKDKEAHREYQTIPVSRM	335	Qy	89	SSLPDTIYHLDGYPEGGIDFTPHNGTKKINTVAYENKLDASGSSTSHLTNNALVEIH	148
Db	301	IVNRYAPIHLKEYMLPTGELLTMDPSNGGHSGTMQRIGEYLHSRQIDNANYGLNSTA	360	Db	123	SSLPTITYHLDGYPEGGIDFTPHNGTKKINTVAYENKLDASGSSTSHLTNNALVEIH	182
Qy	336	IVNRYAPIHLKEYMLPTGELLTMDPSNGGHSGTMQRIGEYLHSRQIDNANYGLNSTA	395	Qy	149	TANGRWREIYLQPGDLEGKAVFREVSSAGCYSSTTFYGDKVTLFSVGNTLFLFKYNGOWF	208
Db	361	LGGENSHSPVVAQLAHHNSRGNYTANGIOHGGGGGGIVTLISTLGNESHEVGHNYGLG	420	Db	183	TANGRWREIYLQPGDLEGKAVFREVSSAGCYSSTTFYGDKVTLFSVGNTLFLFKYNGOWF	242
Qy	396	LGGENSHSPVVAQLAHHNSRGNYTANGIOHGGGGGGIVTLISTLGNESHEVGHNYGLG	455	Qy	209	RSGELENNTITYAQUIWSEELPAHWTIVPGLNLVKQONLSGRNLNDIKIGAPGELLHTID	268
Db	421	HYDGFKGSVHRSLENNNSTMGTDGDKRKFIRPNFYSQTNENKSCLNNOQQEPFDGKFGF	480	Db	243	RSGELENNTITYAQUIWSEELPAHWTIVPGLNLVKQONLSGRNLNDIKIGAPGELLHTID	302
Qy	456	HYDGFKGSVHRSLENNNSTMGTDGDKRKFIRPNFYSQTNENKSCLNNOQQEPFDGKFGF	515	Qy	269	IGDUTTPRDFEAKDKEAHREYFOTIPVSRLMVNTYAPLHKEYVLPTGEELTIDPGN	328
Db	481	DAMAGGSPPSAANRFRTATYPNSAIIORFFENKAVIDSRSSIGFSKWNADQEMEPYHT	540	Db	303	IGDUTTPRDFEAKDKEAHREYFOTIPVSRLMVNTYAPLHKEYVLPTGEELTIDPGN	362
Qy	516	DAMAGGSPPSAANRFRTATYPNSAIIORFFENKAVIDSRSSIGFSKWNADQEMEPYHT	575	Qy	329	GGWSGTGMPQRIGKELYSHIDNANTGUNSTAGLGENSHPYVQAQLAHHNSRGNTANGIQ	388
Db	541	IDRAEQITASVNELESKMAELMAEYAVVKHNGWNTRNYYIPTASADNRGSILTINH	600	Db	363	GGWSGTGMPQRIGKELYSHIDNANTGUNSTAGLGENSHPYVQAQLAHHNSRGNTANGIQ	422
Qy	576	IDRAEQITASVNELESKMAELMAEYAVVKHNGWNTRNYYIPTASADNRGSILTINH	635	Qy	389	VHGSGGGGIVTLQDSTLGENFSEHEVGHYQGLHYDGFKGSKYHRSAENNNTWGMQDKK	448
Db	601	EAGINSYLFINGDEKVVSGCYKSFVDSDQNSUDNDCQLOVDTKQCLRPLRLANHRA NTVMN	660	Db	423	VHGSGGGGIVTLQDSTLGENFSEHEVGHYQGLHYDGFKGSKYHRSAENNNTWGMQDKK	482
Qy	636	EAGINSYLFINGDEKVVSGCYKSFVDSDQNSUDNDCQLOVDTKQCLRPLRLANHRA NTVMN	695	Qy	449	RFIPFYPSPOTNEKSCLNNOQCPFDGHFGFDMAGGSPFSAANRETMTYPTNSAIIQR	508
Db	661	DPEGTLSSTYPA MYGAVFTYSDQNSUDNDCQLOVDTKQCLRPLRLANHRA NTVMN	720	Db	483	RFIPFYPSPOTNEKSCLNNOQCPFDGHFGFDMAGGSPFSAANRETMTYPTNSAIIQR	542
Qy	696	DPEGTLSSTYPA MYGAVFTYSDQNSUDNDCQLOVDTKQCLRPLRLANHRA NTVMN	755	Qy	509	FFENKAVF---DSRSSTGFS---KWN	528
Db	721	KFHINVPTEQPTQATLVCNNKILDKTSUTPAPGLTYTNGOALPKANECCIVSNSG	780	Db	543	FFENKAVSIAVPPASASCMQIRKWN	569
Qy	756	KFHINVPTEQPTQATLVCNNKILDKTSUTPAPGLTYTNGOALPKANECCIVSNSG	815	RESULT 3			
Db	781	KRYCLPVGQSRGSYSLPDWIVQEYVVDSCAKAVLSDNDLYSNRIGEFVGNVPADM	840	TO9438			
Qy	816	KRYCLPVGQSRGSYSLPDWIVQEYVVDSCAKAVLSDNDLYSNRIGEFVGNVPADM	875	Db		toxR activated lipoprotein - Vibrio cholerae	
Db	841	KVKA NGQYLDFSKPRSMRYK 863		Qy		N; Alternative names: toxR activated gene A protein	
Qy	876	KVKA NGQYLDFSKPRSMRYK 898		Db		C; Species: Vibrio cholerae	
Db	876	KVKA NGQYLDFSKPRSMRYK 898		Qy		C; Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 *text_change 17-Nov-2000	
Qy	841	KVKA NGQYLDFSKPRSMRYK 863		Db		C; Accession: T09438; PID:930108	
Db	876	KVKA NGQYLDFSKPRSMRYK 898		Qy		R; Karaolis, D. K. R.; Johnson, J. A.; Bailey, C. C.; Boedeker, E. C.; Reeves, J. B.; Reeves, P.; Parsot, C.; Taxman, E.; Mekalanos, J. J.	
Qy	841	KVKA NGQYLDFSKPRSMRYK 863		Db		Proc. Natl. Acad. Sci. U.S.A. 95, 3139-3139, 1998	
Db	876	KVKA NGQYLDFSKPRSMRYK 898		Qy		A; Title: ToxR regulates the production of lipoproteins and the expression of serum re	
Qy	841	KVKA NGQYLDFSKPRSMRYK 863		Db		A; Reference number: A39108; PMID:91156664; PMID:2000374	
Db	876	KVKA NGQYLDFSKPRSMRYK 898		Qy		A; Accession: A36672; PMID:98165059; PMID:9501228	
Qy	841	KVKA NGQYLDFSKPRSMRYK 863		Db		A; Status: preliminary	
Db	876	KVKA NGQYLDFSKPRSMRYK 898		Qy		A; Molecule type: DNA	
Qy	841	KVKA NGQYLDFSKPRSMRYK 863		Db		A; Residues: 1-1002 <KAB>	
Db	876	KVKA NGQYLDFSKPRSMRYK 898		Qy		A; Cross-references: EMBL:AF034434; NID:93004923; PIDN: AAC12274.1; PID:93004926	
Qy	841	KVKA NGQYLDFSKPRSMRYK 863		Db		A; Experimental source: strain N1661	
Db	876	KVKA NGQYLDFSKPRSMRYK 898		Qy		R; Makino, K.; Ishii, K.; Yasunaga, T.; Hattori, M.; Yokoyama, H. C.; Kubota, S.; Shinagawa, H.	
Qy	841	KVKA NGQYLDFSKPRSMRYK 863		Db		Proc. Natl. Acad. Sci. U.S.A. 88, 1641-1645, 1991	
Db	876	KVKA NGQYLDFSKPRSMRYK 898		Qy		A; Ref ID: Res. 5_19_1998	
Qy	841	KVKA NGQYLDFSKPRSMRYK 863		Db		A; Title: Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemor	
Db	876	KVKA NGQYLDFSKPRSMRYK 898		Qy		A; Accession: T00316	
Qy	841	KVKA NGQYLDFSKPRSMRYK 863		Db		A; Status: preliminary; translated from GB/EMBL/DBJ	
Db	876	KVKA NGQYLDFSKPRSMRYK 898		Qy		A; Molecule type: DNA	
Qy	841	KVKA NGQYLDFSKPRSMRYK 863		Db		A; Residues: 1-587 <KAB>	
Db	876	KVKA NGQYLDFSKPRSMRYK 898		Qy		A; Cross-references: GB:M60658	
Qy	841	KVKA NGQYLDFSKPRSMRYK 863		Db		C; Genetics:	
Db	876	KVKA NGQYLDFSKPRSMRYK 898		Qy		A; Gene: tagA	
Qy	841	KVKA NGQYLDFSKPRSMRYK 863		Db		C; Keywords: lipid binding; lipoprotein	
Db	876	KVKA NGQYLDFSKPRSMRYK 898		Qy		A; Status: preliminary; translated from GB/EMBL/DBJ	
Qy	841	KVKA NGQYLDFSKPRSMRYK 863		Db		A; Residues: 1-25 <PRP>	
Db	876	KVKA NGQYLDFSKPRSMRYK 898		Qy		A; Cross-references: GB:91156664	
Qy	841	KVKA NGQYLDFSKPRSMRYK 863		Db		C; Genetics:	
Db	876	KVKA NGQYLDFSKPRSMRYK 898		Qy		A; Gene: tagA	
Qy	841	KVKA NGQYLDFSKPRSMRYK 863		Db		C; Note: part of the pathogenicity island (VPI); associated with epidemic and pandemic	
Db	876	KVKA NGQYLDFSKPRSMRYK 898		Qy		C; Keywords: VPI	
Qy	841	KVKA NGQYLDFSKPRSMRYK 863		Db		Query Match 15.0%; Score 689.5; DB 2; Length 1002;	
Db	876	KVKA NGQYLDFSKPRSMRYK 898		Qy		Best Local Similarity 24.2%; Pred. No. 9_3e-37;	
Qy	841	KVKA NGQYLDFSKPRSMRYK 863		Db		Mismatches 279; Indels 295; Gaps 39;	
Db	876	KVKA NGQYLDFSKPRSMRYK 898		Qy		4 NSAIYENTSQPINLQGSLAEVKAQSQIQLPAHPQSOPHLSURSLLLVRP-VKA 62	
Qy	841	KVKA NGQYLDFSKPRSMRYK 863		Db		137 NEIDIYHTELEKNTYGSQEGEVTVQTWVII-SPEGRNEPELITGRDLILFKPSIKN 194	
Db	876	KVKA NGQYLDFSKPRSMRYK 898		Qy		63 DDKTPVQYEAADDNNKILGTLTLYPPSSLPDTIYHLDGVPEGGIDFTPHNGTKKINTVY 122	

RESULT 4

BB82276  
ToxR-act  
Species: C. Date:  
Accession #: Heidel  
Heidelberg, R.R.;  
Signature 4  
Title: Reference

Qy	810 AKAKVLLSDPDNLNS-YNRGEFYGVNVNPADMKKYKAWNGOYL--DFS-KPNSMRVY 861
C;Species:	Vibrio cholerae
C;Date:	18-Aug-2000 *sequence_revision 20-Aug-2000 *text_change 02-Feb-2001
C;Accession:	A82494
TagA-related protein VCA0148 [imported] - Vibrio cholerae (strain N16961 serogroup O1)	
Qy	R;Heidelberg, J.P.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Dodson, R.J.; chardson, D.; Ermolaeva, M.D.; Yamatheyen, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P. L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Db	Nature 406, 477-483, 2000
A;Title:	DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number:	A82035; MUID:20406833; PMID:1092301
A;Accession:	A82494
A;Status:	preliminary
A;Molecule type:	DNA
A;Residues:	1-1335 <HEID>
A;Cross-references:	GB:AE004356; GB:AE003853; NID:98657536; PIDN:AAF96061.1; GSPDB:GN001
A;Experimental source:	serogroup O1; strain N16961; biotype El Tor
C;Genetics:	
A;Gene:	VCA0148
A;Map position:	2
Query	Match 14.6%; Score 671; DB 2; Length 1335;
Best Local Similarity 25.5%; Pred. No. 2, 3e-35;	
Matches 231; Conservative 108; Mismatches 249; Indels 318; Gaps 36;	
Db	16 NDLOGSLSAAEVKAFOSQILPAHPPK--EGDSQPHPLTSL--RKSLLLYRVPVKADDKTPVQE 71
353 NHQSLSLEGGLSISIQTHTSYA-PAGNELTGQHDAIMNREALLETPQOGEEINOVRAE 411	
Qy	72 ARDDDNKILGLTLYTYPSSSUPDTYHLQDGYPEEGIDTPHNGTKKIINTVAVNKLSDAS 131
Db	412 VFLDGELYQFTMLL_PPSA-----ASDOPENGRMKV ----- 443
Qy	132 GSSITHSHLNNNALVEIHTANGRWWDIXLPQGPDPLEGKMRVFSAGYSSTVYGDHKVT 191
Db	444 --VSHLA----- 449
Qy	192 LSGVNTLLFKYVNQWFRSELENNRITYAQHISWALPAHNTVPGNLNVI---KQGN 246
Db	450 ----- 450
Qy	72 ARDDDNKILGLTLYTYPSSSUPDTYHLQDGYPEEGIDTPHNGTKKIINTVAVNKLSDAS 131
Db	451 WSPLQLQDWKPKGSLRLRDNLGREGV 476
Qy	247 LSGRLNDIKIGAPGELLHTIDIGLMATTPDRDFEADK-KEAHREYFOTIPVSRMIVNNY 305
Db	477 LS--OGEIOFGAPELVONIDICHLMPPRDRNMQIQLPTLAADYOKIPASKLYHADY 534
Qy	306 APLHLKEVWLPPTGTGELJTDMDPGNGGWHSCTMRQFLGKLYSHGIDNANYGLNSTAGLGEN 365
Db	535 TPAHFPTVMTPTGGVYTYTDSASTASGWSHCDMEKAGKANVSYINNANVGYSSAGSYSS 594
Qy	366 ----SHPYVVAQLAAHNSRGNTAN-----GLOVRHGSGGGGIVLTDLSTLGNEFSHEVG 414
Db	595 YNRRNHH-----ITAHTHNGTYITKKDTDLPOVYVHGGGGGIVTLEATIGNEWSHELG 648
Qy	415 HNYGLGHYDGFKGSVHRSAAENNNTWGMGDGDKRFLPNF----YPSQTMNEKSCLNNO 468
Db	649 HNYGLGHW--PYMASLH----DLSLGWGMDAFHGRFLGNLHWGDTYTQQQGD-----D 696
Qy	469 CQEPP-DGHKEFGFDMAAGGSP--FSAANRFTMYTPNNSAIQRFENKAVFDSRSSTGFS 525
Db	697 IYPPPFDAFREFLDRQDQNGGEEQYGTISRTELEPAQRERKAQWMNNNSFLDSHSPCGYY 756
Qy	526 KWNADTOQMEMPYETIDRAEQI----- 557
Db	757 QWDQETQRYKAVETDTPKPOQVGPPVTLGLIDQPQNENPSOIYPLTSNYGNFEQPQP 816
Qy	558 KMAELMAEYAVVHMGWNTWRYIPIASADNRCGSLTINHEAGNSY--LFINDEK 615

A;Cross-references: GB:AE001437; PIDN:AAK80209;1: PID:Q15025253; GSPDB:GN00168  
 A;Experimental source: Clostridium acetobutylicum ATCC24  
 C;Genetics:

A;Gene: CAC2252	Query Match Score 3.6%; Best Local Similarity 18.6%; Matches 179; Conservative 121; Mismatches 319; Indels 343; Gaps 48;	DB 786 PV 787 Db 876 PI 877	Qy 786 PV 787 Db 876 PI 877
RESULT 8			
56 LYRPVKA---DDKTPVQEARDNNKKLIGLTLYPPSSLP----DTIYLDGVPEGGI 106	Qy 56 LYRPVKA---DDKTPVQEARDNNKKLIGLTLYPPSSLP----DTIYLDGVPEGGI 106 Db 29 TIRPSAKAPRMTISKTKTKEYAR----SNSLAKLNGDILQYNGLDETDI 79 Qy 107 DFTPHNGTKKLINTVAEVKNSDAGSSSIHSHLTNNALV----EIHTA 150 Db 80 KICBQVLWYKDP-SGQSSDPLVVDENKINN-TGNLISLSDNPDAVITORMTIKIS 137 Qy 151 NGFWYRDIYLPQGPDLGKMW--RFYSSAGGSSTVEYGDRTVTLWSGNTLFLKYNGW 207 Db 138 KSDLTMSVY----DSTGRQIVKQCIASRSKVSFTHNSGDR----- Qy 208 FR---SGELEN-NRITYAQHNSAELPAHWYVPGNLVYKQNLGRBLND----- Db 182 FKEDSNKGMRNGTESVTAGYQCHCGSFPEWNSDGYGLVSDGGSPFTIGTSLOYSGIS 241 Qy 254 -----TKIYGAPELLHTIDI-----GLHATP----RDRFDFAKDEAHRE 290 Db 242 KTDIYDYLUGSPREVISESDYSGSKAPMFKPATGFTNTQWNNNSLSGTGNDKLS 301 Qy 291 YFQT----IPVSRMIV----NNTYAPLHLKEYALP----TGELLTDMDPGNGGWHS 333 Db 302 VLNTYRSKQIPDNFCCLDFERFKKNGDNEYGEFKNTNDPDAQNLKAYMD-SKGKLM 360 Qy 334 GTMRORI----GRELVH----IDNYANGLNSTAGLGENSESHPYV 370 Db 361 GIMDKPLADSEQARYVTSGWWLQGDSAAIDSCKRAMEVNAFISDV----- Qy 371 VAQLAAHNSRGNYANGIQYHGGSGGGGIVTYLDSTLGNFESHEYGHNTGLGHY 422 Db 410 -----RKKWNNNIQ----DAFDKSGIVGFWNDECDENVNFGFNMMERA 450 Qy 423 -VDGFKGSVHSAAENNNTWIG-----WDGDKKRIFPNFYSOTNEK-- 462 Db 451 IYDG----QRAYKNORVWSLNRRNYAGAQRYCYGMNSGSD----ISTGFDSMANQERM 500 Qy 463 -CLNNNOQCQEPFDGKIFGFD -AMAGGSP ---ESAANFTPTMTPNSAIIQFFENKA 515 Db 501 LSAVN----LGEAKWGMETGFGNDGDPPTENPARMEFSAFATP----IFRHYQGDNRV 550 Qy 516 ----FDSRSSTGFSKWNADTOEMEPYHTIDRAEQITA-----SYNLESESK 558 Db 551 RYPWAGGSTAZAAKAMQDRTPLIPIYITSYDRAQSOGSLVSLMEYPRDSSNAANDK 610 Qy 559 MAELMAEAYVTKVHMNGNNTRNVIY-----TASADNRGSLTINHEAGNSY---- Db 611 EAWMFQDMYLSPVNVQGQTSKSIY-PEGWIDYTGREYTTGG-OTINVAVSTDNWSDLP 669 Qy 608 LFING-----DEKVVSCGCKKSFVSD-----QFWKER 635 Db 670 LFIKSGAIITPDQFENYVGKEKITDVYDFPSDKATTFFYDDGTSYDENGNSYFDOX 729 Qy 636 DVVDTREARKPEQFGVPTVLYKDPEGTLLSSYI-----YP 672 Db 730 MFLQTSDDSKVQFNIDKNT---GSYTPD-LKDYVYKMHYKGAVTANGQALTOYSSYD 785 Qy 733 TOATLVCNNK----LIDTKSUTPAREGUT -YTYNGQALPAKENEGCIVSVNSGKRYCL 785 Db 827 VITAAAANPKGGTYYGPOTVSLTAKSDATIYTTL-----DGTTPTVNSTK-YTA 875 Qy 845 GSPSPFSANRPTMTPNSAIIQRFENKA VFDSSRTGISK----WNAQTOQMEMPYERI 541 Db 827 GEE-----QHFHRN-----GFEQQGWDSTYQE----P1DA 566	Qy 16 NDLOSSLAAEVKFQASQSLPQAPKREGDSOPHLTSURKSSLVVR----PVKAADDKTP--v 68 Db 131 NDMTGSMVMA-----IKDELNSKG-----ITNAQDMLSGIAQSVTISNPTPGGGA 177 Qy 69 QYEARDDNKKLGLTLYPPSSLQDPTYHLDGVPEGGLDFTPHNGTKKLINTVAYEVKK-- 126 Db 178 QIRTRGSSL-----NASNDPLVIVDGL---AID---NEGIKCHANGLSMVNPAD 221 Qy 127 -----LSDAGGSSIHSHLANNALVEHTANGRWYRDLILPQGDLEGQMVRETVSSAYS 180 Db 222 IETLTULKDSATAIYGSRSASNGVIIITKKKG-----NGQAPSNTVNGSYFS 270 Qy 181 STVFGDRKVTLSVGNNTLFLPKYNGCWFNSRGELENNRITYAOHWSAELPA-----HW 233 Db 271 KT-----OKRDVLSGDEYRA-----YANLQNGDLPADLGXTNDW 3C7 Qy 234 -----IVPGLNLVTKOG --NTSGRLNDIKIGAPELJLHTIDGMLTTPRDRFDFAK 283 Db 308 QDQIFRAVSTDHVHSINGGFKNLRYRS-----LGYTDONGTKTSNFR-RITA 356 Qy 284 DKEAHRYFQ---TIPVSRMFV---NNYAPLHLKE---VMLPREGELTDMDP---GNGG 330 Db 357 SYNLAPSFFEDHLKFNNNAKFMNGKNRYADSRCSRYTRALAIDTPRVYNSNEDPYQFTGXY 416 Qy 382 NYANGIOYHGGGGGGVTLDSLGNFESHEVGHNYGL-----GHYVDGFKG 428 Db 455 N-----NSHPYTVQAQLAHHNSRG 381 Qy 417 WQN-----INSTIGFSNPDWKTSNPNSPQNPLAELELKNOKG 454 Qy 429 SVRSAEENNNTWGMODD---KKRFLPNTYPSQTNKESCLNNNCQCEPFDGHKFGFDAMA 484 Db 492 TIVSPYSGNNYGRNGDVTQKYNLYSYIYQI-KSLGAND-----FDLNV 538	Qy 331 WHSGTMFQRIGKELYSHGIDNANYGLNSTAGLG-----NSHPYTVQAQLAHHNSRG 381 Db 417 WQN-----INSTIGFSNPDWKTSNPNSPQNPLAELELKNOKG 454 Qy 429 SVRSAEENNNTWGMODD---KKRFLPNTYPSQTNKESCLNNNCQCEPFDGHKFGFDAMA 484 Db 492 TIVSPYSGNNYGRNGDVTQKYNLYSYIYQI-KSLGAND-----FDLNV 538

QY	542 DRAEQITASVNESESKMA---ELMAEAYAVVKHAMNGNNTRNIYIPTASADNRGSILT 597	DRKVTLVSGNT 197
Db	567 KLRQQTAYATRNLTVSYGRLLNNSLNNYLFLFTMRDGS-----SRESDNWR---- 615	1597 ---NFLQSNIKPFILDANFIISGTTSIGQQFFICDENNNIOPYFIFKFTNLTNTLYGMR 1653
Qy	598 INHEAGAQNNSYLEINGDEVYSSQGKKSKFSVSDQWKFKEBDVDTREAKPQECPVPUTLV 557	198 L-LEKYYNGQWFRSGELENNRITAQ-HIWSAELPAHWWYPGLNLVWKGNLNSG----- 249
Db	616 -----GTFSSLALQWK---IEKENFLKVNVLSDLKR---LGWGTGOO 654	1654 QNMIVEPNVDIDDDSDCISSTVINFNSQKYLIGIDSCKVNKVVISQVTKRFVNFKY 1713
Qy	658 GYDPGETLSSXYYIPAMYGAQF-TYSDDS---QNLSDNDCCQLQVDTKEGQRLERBLANHR 713	250 -----RLNDIKI-----GAPGELLHTIDIGMELTPDREF-DEAK 283
Db	655 NIGDFDFAVPLYYVNNNEAQYQPGDGTYSTSRPKAFENLYJEKTIIWNAGDGEFLNGR 714	1714 TYPEIVLVDANYINERKINYNINDSLRTYWSNDGNDFINLMSTSEENKVQSQVTKRFVNFKY 1773
Db	714 ANNTVWANKPHINVPTESOPTAVLVCNNKLIDTKSLPAPERLY-----TVNGO 763	284 DKE-AHREYF---QTIPVPSRMVNNYAPHLAEVML----- 315
Qy	715 ITGGSDGYF-----RKIMTCYTALESPPMNILQCPDTEYRETGKLFYGFSTNAK 763	1774 DKTFLANKLSPNFSDRODYPSEIIL-SETPSYEDGLIGTDGLGLSYLYNEKFYINNFQHM 1832
Db	764 ALPAKE-----NEGCTIVSYNSG-KRYCLPYG----- 788	316 -----PTGEELTD-----MDPGNGWHSGTMORRIGKELYSHG 349
Qy	764 PIYTKDFPTWDLSSYNITWTHNEITKLTGDDSDSYVEAGDKISIERNNTKVKQAHKGVAANS 823	Db 1833 VSGLIXYNDISLIIYFPPVNLITGEVTYGDKYKFNPINGAAS-----IGETI---- 1882
Db	789 -----QRSGYSLPDMTGVQEGYVYDVSQGAKAVLISDWDNLSSYNRIGEFYGNVN 835	Qy 350 DNANYGLNSTAGLGENSHPTVVAOLAAHHSRGNV----ANGIQYHGGGGGGITVTLDSL 404
Qy	824 FIVYSGRNNTKVKQAHKGVAANSFVQYQVYDENGKPI-----ENHFVDRGN--GTID 874	Db 1883 DDKKNTYFNQSG-----VLTQGVFVSTEDGKFYFAPAN-----TLDEN 1918
Dbb	836 PADMKKVKAWNG-----QYLDPSKPRSMR 859	Qy 405 LGNE---FSHEVGHNYGLAHYVDGFKGSVIRSAENNNSTWGW-----DGDKRKFIFNFYPPSQ 453
Qy	875 SGDKIYKPKPGDYLMGLTSKIQKYNFDFSFSLR 908	Db 1919 LEGEADTFGTGLIIDIINNYYFEDNTRGAVE-----WKEDTAGHYYFSPE---- 1962
Db	RESULT 9	Qy 459 TNEKSLCNQNCQEPFDGHKFGFDAMAGGSPFSAARF-TMYTPNSSAIIQRRF---EN 512
Db	S70172	Db 1963 -----TGRAFKGLNQIGDKYYFNSDGTMQKGEVNSINDN 1996
C;Species: Clostridium difficile	Qy 513 KAVFDSRS---STGFSK-----WNADTOKEPEYHTIDAEQTAVSNELSESQMAELMA 564	
C;Accession: S70172; S44271	Db 1997 KHYFDDSGVMKVYGYTEDGKHFYFAENGQVNTEDGFKYFAHHNEDLGNEEGEELS 2056	
Qy	565 BYAVVKVHMNGNNTWNTIYPTASADNRGSLLTINHEAGGINSYFLINGKVKVSSQYKKS 624	
Db	2057 YSGIL-----NFNNKYY---FDDDSFTAVVGKNDLEGDSKYTF---DEDTAGYIGLS 2103	
Qy	625 FVSDGQFWKERDVDTREARKPEQFG-VPTTLYGYPDPEGTLLSSYIYPAAMYGAQGFTY 683	
Db	2104 LINDGOYFNDGJ-----MOYGFVTLNDKVEYFISDGII-----GVQNI 2145	
Qy	684 DDSQNLSNDCQLOV---DTKEGQLRFLRLANHRANTWVNKFHINVPTESOPTATLVCN 740	
Db	2114 DDDNYFYIDDSNQVQIGVFDTSQGKYFAPA-----NTVNDNTY----- 2183	
Qy	741 NKLDKTSKLTAPE---GTYTQVYDQALPAKENEGCIVSNSAKRCKYLPVGORS---GX 793	
Db	2184 GOAVEYSGLVRVGEDVYYFGETYTIETGWYIYDME-----SKYFYPETKACKGI 2246	
Qy	794 SUPDWIVQGEYVVDGAKAKYKLLSDMDNLSY--NRIGEF-VGNNPNPAD 838	
Db	2237 NLIDDI---KYPFDERGIAARTGLISFENNNYFNGENIQFGYINIED 2281	
RESULT 10		
F97255		
C;Species: Clostridium acetobutylicum		
C;Date: 14-Sep-2001 *sequence_revision 14-Sep-2001 #text_change 14-Sep-2001		
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Daly, M.J.; Bennett, G.N.; Koontin, E.V.; Smith, D.R.		
J;Bacteriol. 183, 4823-4838, 2001		
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium		
A;Molecule type: DNA		
A;Cross-references: CPB-A5001437; PMID:21359325		
A;Status: Preliminary		
A;Residues: 1-1157 <RTR>		
A;Cross-references: CPB-A5001437; PMID:21359325		
Query Match 3.5%; Score 159.5; DB 2; Length 2367;		
Best Local Similarity 18.4%; Pred. No. 0.14;		
Matches 188; Conservative 149; Mismatches 321; Indels 353; Gaps 49;		
4 NSAIYFNTSOPINDLQGSLAAEVKFAQSOLLAIPAHPKEDGSQPHLTSLRKSLLVRPVKA 63		
1454 NSELQKNIPYSFVSEGK---ENFGINSTKEG---LTVFSELPDVFLISKYMD 1501		
64 DKTP-----VOVERDANNKILG-----TLYTPESSLPDTIYBLDG 101		
1502 DSKPFFGYYSNLNKDVKVTKDNYNLIQGYLKKDIKSLSLQDEKTIKNSVHD-- 1559		
102 PEGGIDFTPHNGTKIINTVAENKLSDASGSSIHSHLTTNALVALTEITANGKVRVDIYL 161		
1560 -ESG-----VAEILKFMNRKS---TNTSDSMSFESMN--IKSIFV- 1596		

A: Experimental source: Clostridium acetobutylicum ATCC824  
C: Genetics:

A; Experimental S  
C; Genetics:

Db	392 NVILLEQPKITVHKYVGIMSTEENKNEQQ--DNTNTIGL-----AKMIA 434	Db	1453 NSELQKNPYSFVDFSEKG-----ENGIFINGSTREG---LEVSELPDVLLISKYMD 1500
Qy	377 HNSRGNYANGIQHV-----	Qy	64 DKTP-----VQEARDDNKKILG-----TLTLYPPSSLPDTIYHLD-- 99
Db	435 LNSKNEYKELESHKRAWTYDLYNDAAFILEIPSDSILEMARTSLEFHLLANTRDYNVSSDRG 494	Db	1501 DSKPSRTGYISNNLKVTKVITKDVKVNLITGYLKDDIKISLSTLQDETIKLNSVHLD 1560
Qy	391 ---GGSG----GGCIVTLDSTLGN-----FSHEVGHNYGLGHYDGFKGSSVHRSANN 437	Qy	100 GYPEGGIDTFPHNGTKKINTVAEYNKLSDASGSSTIHSHTUNNALVEHTANGRWRD- 157
Db	495 LPVGYSGLSSDSYGMVFADWMEPAUPLPFVNVAQN-----MHNRYRNATHSQAKLN 548	Db	1561 GVAE-----ILKFHARKRGNTN-TPSLMSPLESENKIKSFVNFLQSNTKFILDAN 1609
Qy	438 NSTWGDGDKRFKRFIPNFYPSQTNEKSCLNNCQEPFDGKFGFEDAMAGGS----- 487	Qy	158 -----YIYPOGPDLEGDAVRYTFVSSAGYSSTYFYGDKVTL-----S 193
Db	549 AEKYGYGAAYPWTSKQYANCTSTGCPVDEYEHINVDYMASFSIYLNGHEGIDDEYLR 608	Db	1610 FIISGTTSIGOFEFICDENTOPIYPFKF-NTELNTNTLVIGNEQNMIVTEPYNDLDDSGD 1668
Qy	488 ---PF--SAARNTMTP-NSAATIQRPF-----NAKAVDSSSTGFSKKNAD 530	Qy	194 VGNLTLL--FKYNGOWFRCSELEN--NRTYAQHWSAEI--PAH-----WIVPG 237
Db	609 TWPLTKNAQEFATAYVKVNSLGIYETINLTDPDEFANHINNGAFTNAGITKLKWD 668	Db	1669 ISSTVINFSSQRLYG-----IDSCVNVVYSPVITYEINITPVYENNTNPPEVIVLD 1721
Qy	531 TQEMPEYHFTIDRAQITASVNEUSESKHAELMAEYAVYKHMNGNTRNVIYPTASAD 590	Qy	238 LNVIKOGNLGRNLNDIKI-----GAPGEELLTFIDIGHSLTTPRDRF-DFARDKE-AHR 289
Db	669 -----IGNHLEGVVDPKWEI-----SKDLYIPRSSSN 696	Db	1722 ANYNEKINVN--INDLSIRYVWSNDGNDFFLMSTSEEKVSQVKIRFNVFDKTLLAN 1779
Qy	591 NRGSILTINHEAGYNSYL-----FINGDEKUVSOGYKKSFSYSDQGFWKER 635	Qy	290 EYF-----QTIPVPSRMVYNNYAPHLKEYML----- 315
Db	697 -----ITLEY-SGMSSVETIKQADVTLMVPLGYIN-DESILNAIK-----DLVYFSE 744	Db	1780 LSFNFSDKDQDVPSVETL-SETPSYIEGGLIGDGLVSLVLYNEKFYINNGMAYSGLIY 1838
Qy	636 DVVDTRAREAPEQ--TGVPVTTLGYDYYOPESTLSSXYTAMYXGAYGFTYSSDDSONLSDN 692	Qy	316 -----PRGELLTD-----MDTGTGNGGWHSCTPRQRIGKELYSHGDIDNANTGGL 356
Db	745 -----OSASGFMAMTIPVFEVAAAAGLNLNGSSQSLYKSVPLRAPPAQQFSEQDDN 797	Db	1839 NDSDYYFRPPYVNLITGFTVGDQKYYFNPINGGAS-----IGETI-----IDDKNKF 1888
Qy	S10317	Db	357 NTAGLGENSEHPHYVVAQLAHHNSRGNY-----ANGIQVHGSGGGGIVTLDSTLGN-----408
Db	toxin B - Clostridium difficile	Db	1889 NOSG-----VLTQGVFSTEDGFKYKAPAN-----TLDENLEGEAD 1924
C;Species:	Clostridium difficile	Qy	409 FSHEVGHNYGLGHYDGFKGSSVHRSALENNNTSGW-----DODDKKRPINPFSQTNNEKSCL 465
C;Accession:	S10317; S22444; S2244	Db	1925 FTGKLIDENITYYFDONYGRAYE-----WKELDGEMHESPE----- 1961
R;Barroso, L.A.; Wang, S.Z.; Phelps, C.J.; Johnson, J.L.; Wilkins, T.D.	Nucleic Acids Res. 18, 4004, 1990	Qy	466 NNOCQEPFDGHKFGFDAMAGSFSPTSAANRFF-----ENKAVFDSR 519
A;Title:	Nucleotide sequence of Clostridium difficile toxin B gene.	Db	1962 -----TGKAFKGLNQIGDQKYYFNSDGVQMKGFVSINDNRHYFDDS 2002
A;Reference number:	S10317; MUID:90326540; PMID:2374729	Qy	520 S--STGFSK-----WNADTOQMPEYHTIDRAEQITASYNLESKSMALMAYAVVK 571
A;Accession:	S10317	Db	2003 GMRVKGTEIDQKHFIAERGMQIGVPTNTEDGFKFAHNRDNEGEEIGTSGIL-- 2060
A;Status:	translational not shown	Qy	572 HMNGNNTNRTNIYIPTASADNRGSLTINHEAGYNSYLINGDEKVYSGYKKSFSVSDGQF 631
A;Molecule type:	DNA	Db	2061 -----FDDTAETIGLSLINDGQY-----NPNKNTY-- 2109
A;Cross-references:	EMBL:X53138; PID:G40442; PID:CAA37298; PID:G40443	Qy	632 WKERDVDTREARKPFGF-YVPTTLYGYYDPEGLTSSYIYPANYGAYGFTYSSDQNL 690
R;Eichel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerboe	Mol. Gen. Genet. 233, 260-268, 1992	Db	2110 YFNDDG1-----AQYGFVTDKVFYFSDSGIES-----GVONIDDNYFYI 2151
A;Description:	Comparative analysis of Clostridium difficile toxins A and B.	Qy	691 DNDTCQLQY-----DJKEGOLRFLRLAHNTANTVANKHINVPTESQPTQATLYCNKKILDTK 747
A;Reference number:	S22434; MUID:92293124; PMID:1603068	Db	2152 DBNGTVQGVFTSDQSKYFPA-----NTYONDY-----GQAVEY 2189
A;Accession:	S21894	Qy	748 SLTPAPE-----GLTYYNGOALPAKENEGCTIVSYNSGKRYCLPYGQRS--GYSLPDWTV 800
A;Gene:	toxB	Db	2190 GLYVRGEDEVYFGETITETGWMYDNE-----SDKYFNPETKACKGINLDDI- 2241
C;Superfamily:	cpl repeat homology	Qy	801 QGYVYVDSGAKVJLSDWDNUY--NRIGEF-VGNNVPAD 838
C;Keywords:	cytotoxin	Db	2242 --KYYFDEKGTRGTLISFENNNYVNGENQFGYNTIED 2280

RESULT 13  
AB3412  
hypothetical protein PA1B74 [imported] - *Pseudomonas aeruginosa*  
C;Species: *Pseudomonas aeruginosa*  
C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

Query Match 3.3%; Score 153.5; DB 2; Length 2366;  
Best Local Similarity 18.6%; Pred No. 0.35; Gaps 51;  
Matches 186; Conservative 155; Mismatches 321; Indels 339;

4 NSAIYFNTSQPINLDQGSLLAEYKFAQSQILPAHPEKGDSOPHLSLRLKSLLYRPyKAD 63



Qy 456 PSOTNEKSCLANOCPEPDGHF DAMAGSSPFSAANRTMVT -PNSSAIIQREFFENKA 514  
 Db 575 IRLDSSNANLKYQDHQER DNYENLT -WPFVADDYLVRHAKNPNRIVSESEVAK 632  
 Qy 515 VFDSSRGTSFGSKWNADTOEMPEYHTIDRAEQT -ASVNDESESMAELMAEYAVVKV- 571  
 Db 633 LKDASEKVKSGO-----YFSFTLRSSTGASKGLDQFAMNKFFDDSLKKLD 682  
 Qy 572 -HMGNGWTRNVIYIPTASADNRGSILTINHEAGNSYLEINGDEK-----VVSQYKKS 624  
 Db 683 TYSWSGCKTLLTAFTNTYKVDSNN---RVTYDVFHGTLPNEGDSKNSLSPYGRKINGTYKCT 739  
 Qy 625 -----FVSDGQE-----WKEKD-----VVD----- 639  
 Db 740 EKEKIKESESEGSDPDKIVSYEWFEDGDNKSNNEENPEHSYDQVGTYTVKVTDDKGES 799  
 Qy 640 -----REXARKPQFGVPTTL-----VGTDPGETLSSVTPAMYGAY 678  
 Db 800 SVSTTTAEKDLSENKLPVYHVPKGALNQKVTFYKGKTYDPGSIAGYQWD----F 854  
 Qy 679 GFTYSDDSQNLSDNDCOLQVDTEGOLFRFLANHRANNTVMKFHNIVPTESOPTQATLV 738  
 Db 855 G-----DSDDFSSEQNPSPHYTKGEVTVL-----RVMD-----SSGQMSKXTM- 894  
 Qy 739 CANNKILDTKSLSLP-----APEGLTYTNGQALPAKENEGCITYSVNSGKRYCILPV--GQ- 789  
 Db 895 --KIKITDLPVYPIGTEKEPNRKETASGPVPGIPUSGTINTSDODYFVTPGEV 951  
 Qy 790 -----RSGYSLPDMWQEVYVDSGAKAKVLLSDWDLNSYNRIGEFVGNVNPADMKKVKA 844  
 Db 952 RIDINKLG3YGGATHV---VYDENNNAVSYATDDGQNL-----GKFKADKPRYTHLYM 1003  
 Qy 845 WNGQYLDL 852  
 Db 1004 FNGSYMPY 1011

RESULT 15  
 AC1333 surface protein (LpxTG motif) [Imported] - *Listeria innocua* (strain Clip11262)  
 C;Species: *Listeria innocua*  
 C;Date: 27-Nov-2001 \*sequence\_revision 27-Nov-2001 \*text\_change 27-Nov-2001  
 R;Accession: AC1333  
 R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, H.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.  
 Science 294, 849-852, 2001  
 A;Authors: Krecft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Weiland, A;  
 A;Title: Comparative genomics of *Listeria* species  
 A;Reference number: AB1077; MUID:21533279; PMID:11679669  
 A;Accession: AC1333  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-1386 <GLA>  
 A;Cross-references: GB:AB592022; PRDN:GAC96035.1; PID:916413254; GSPDB:GN00178  
 A;Experimental source: strain Clip11262  
 C;Genetics:  
 A;Gene: lin08082

Query Match 3.2% Score 147; DB 2; Length 1386;  
 Best Local Similarity 18.9%; Pred. No. 0.4;  
 Matches 166; Conservative 113; Mismatches 297; Indels 300; Gaps 44;

Qy 110 PHNGTKIINVAEYNKLSDA-----SGSS1-----HSHLTNALVEIHT--ANGRWRI 158  
 Db 406 PYNGTNTNPRTKT1 -LSDSLTMKYDGTMGYHWSAVNNIAIDSTIYSSDDWKAED 463  
 Qy 159 YLPOGPDLESKMYRF--VSSAGYSSSTVFGDRKTYLSTVNTLFKYVNGQWFRSGELENN 216  
 Db 464 NFISATOKDPLPLVTVGSDTTPEXEYETSYVNGLTTITVVKENOANIVADES 1187

Search completed: October 8, 2003, 16:28:19  
 Job time : 25.2134 secs